Quantifying Productivity Growth in Medical Care: 20 Years of Evidence From Nine Health Conditions

Authors Calvin Ackley, Abe Dunn, U.S. Bureau of Economic Analysis, and John A. Romley, University of Southern California[1]

Contact calvin.ackley@bea.gov

Date March 2026

Abstract Productivity growth is the fundamental driver of improvements in living standards. However, accurately measuring productivity remains challenging for the health care sector, which accounts for about 18% of the U.S. economy. We develop a framework that combines a utility-based measure of output with a cost-based measure of inputs, consistent with core principles of productivity measurement. The output measure captures improvements in both longevity and quality of life, while the input measure reflects underlying treatment costs rather than regulated prices commonly used in prior studies. We apply the framework to analyze nine conditions over the period 2002–2021, providing the longest and most comprehensive assessment of productivity for acute conditions in the literature. While official productivity measures omit welfare gains from improved health and therefore show flat or declining productivity, our central estimates imply productivity gains of 7.5% per year for these conditions. This result is consistent with the large welfare gains associated with improvements in population health documented in prior research. We find the direction of productivity gains to be robust, although the magnitudes are sensitive to the value of health improvement.

Keywords Productivity, quality-adjusted prices, medical care

JEL Code E31, O47, I10

[1]We would like to thank Sandra Decker, Dennis Fixler, Anne Hall, Steve Hill, Eli Liebman, Giacomo Meille, Sam Zuvekas, and the Agency for Healthcare Research and Quality Seminar participants for helpful comments.

Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE

1. Introduction

Productivity growth is a fundamental driver of improvements in living standards. However, accurately measuring productivity remains challenging for key sectors, including the health care sector, which accounts for about 18% of the economy (Hartman et al. (2026)). Measuring productivity in health care is challenging for two main reasons. First, both the costs and benefits of technological change are difficult to measure. Second, even when outcomes improve, it is often unclear how much of the improvement is attributable to medical care rather than to broader changes in population health.

This paper provides the first framework for measuring productivity in health care that combines a utility-based measure of output with a cost-based measure of inputs, consistent with core principles of productivity theory (Syverson (2011)). Prior research typically addresses only one side of this measurement problem. Studies such as Dauda et al. (2022) and Cutler et al. (1998) estimate utility-based, quality-adjusted prices for acute conditions, improving output measurement, but using observed payments rather than underlying treatment costs. In contrast, other studies, including Romley et al. (2015) and Romley et al. (2019), use cost-based input measures but define output using the quantity of successful treatments rather than the full utility value of improved health.[2] By combining welfare-based output with cost-based inputs, our approach provides a conceptually improved measure of productivity in the health care sector. Moreover, our analysis spans a longer time horizon and a broader set of conditions than prior work.

We apply this framework to nine well-studied health conditions in the Medicare population over the period 2002–2021. In the first step of the methodology, we risk-adjust key variables for each condition. This step ensures that cost and outcome variables can be compared consistently over time. We then estimate productivity using index number methods.[3] Specifically, we use index formulas to compute real output growth, real input growth, and then combine them to estimate productivity.

Relative to the existing literature, our approach introduces three methodological improvements. First, we account for costs, in contrast to prior work that measured payments. We provide empirical evidence that costs and Medicare payments diverge in our sample, with payments growing roughly 0.7% per year more slowly than costs, reinforcing the need to use cost-based input measures when assessing productivity. Second, we account for quality-of-life gains, recognizing that failure to return home or unplanned

[2] The approach taken by Romley et al. (2019) and Romley et al. (2015) presents an approach common in economic measurement that uses a ratio of quality change for quality adjustment. However, as shown in Sheiner and Malinovskaya (2016) and Dauda et al. (2022), output measures based on successful treatments can substantially understate the value of medical care relative to a utility-based approach.

[3] We focus on index methods that have fewer data requirements relative to more structural approaches, such as in Grieco and McDevitt (2017).

readmissions substantially reduce health-related quality of life. Third, we improve the measurement of long-term survival by allowing expected gains to vary with patient age and other factors, potentially allowing for higher long-term life expectancy for younger patients, relative to older patients. In addition to these improvements, we extend prior work by covering a longer period, including the post-2015 ICD-9 to ICD-10 transition.

Prior work by Cutler et al. (1998), Romley et al. (2015), Romley et al. (2020), Matsumoto et al. (2021), and Dauda et al. (2022) also focuses on acute health conditions.[4] An advantage of studying acute conditions is that risk adjustment is relatively well understood. As a result, once we adjust for patient severity, changes in outcomes and costs can be more credibly attributed to medical care rather than to external factors such as patient behavior, underlying health, genetics, or environmental influences. The conditions we study (heart attack, heart failure, pneumonia, gastrointestinal bleeding (GI bleed), fractured hip, stroke, coronary artery bypass graft surgery (CABG), lower-extremity joint replacement (LEJR), and chronic obstructive pulmonary disease (COPD)) cover a substantial share of medically intensive care.[5] While this represents a limited set of conditions, they are economically important, accounting for roughly 10% of Medicare Part A and Part B spending (Romley et al., 2020).

We present evidence of quality improvement over time for all conditions studied. Survival improves, there are fewer unplanned readmissions, and patients are increasingly likely to return home after the 90-day period. We also find that the treatment costs for these conditions have increased faster than economy-wide inflation, but that the value of the quality improvements greatly exceed the cost.

Aggregating over conditions, we estimate large productivity growth of 7.5% per year when the value of a quality-adjusted life year ($VQALY) is set to $100,000. We observe differences in productivity growth across disease conditions, with the largest gains for hip and knee replacements and pneumonia, with smaller productivity gains for heart failure and COPD. As in other studies, estimates are sensitive to the assumptions made regarding the value of a statistical life year, with a productivity measure of 4% per year when the $VQALY is $50,000, and around 10% per year when the $VQALY is $150,000. To put our productivity estimates in perspective of aggregate health sector productivity, we multiply the productivity gains we observe for the select conditions by the share of expenditures those conditions represent in the broader health care sector, which is about 3.8% (based on estimates from the full population from the Bureau of Economic Analysis (BEA) Health Care Satellite Account).[6] We estimate

[4] See Sheiner and Malinovskaya (2016) and Sheiner and Cutler (2024) for a more extensive review of the literature.

[5] With the change in disease coding from 2015 to 2016, we are not able to track GI bleed, stroke and hip fracture applying the same methodology, but we are able to study CABG post-2015. The number of conditions tracked shifts from eight conditions up to 2015 to six conditions in 2016. Lower-extremity joint replacement is not formally an acute condition, but it is similar in that the relevant measurement issues are well understood, permitting a close link between treatment and outcomes.

[6] These conditions are more important for the over 65 population, explaining the greater share of expenditures for

that this implies an aggregate health sector productivity bias of –0.29 percentage points per year.

There are a few limitations and robustness checks worth noting. First, our analysis focuses on Medicare fee-for-service (FFS) patients, who represent a declining share of the Medicare population as Medicare Advantage (MA) has grown. We believe this has a limited effect on our estimates, as focusing on select conditions helps standardize the population being studied. Moreover, by applying risk adjustment to the patients, this further normalizes the sample and captures the changing treatment costs and outcomes consistently. Second, changes in coding practices may raise concerns about potential bias, but we find large productivity gains when we use no diagnosis codes in our risk adjustment, although the magnitude of the productivity gain is lower.[7] Third, our base analysis focuses on a 90-day spending window, as we show that the vast majority of episode-related spending takes place within this period. However, we perform robustness analysis that assumes that the episode may affect lifetime spending. These alternative specifications continue to show sizable productivity gains, although the estimated magnitudes are again somewhat attenuated.

Following Syverson (2011), productivity in the economic literature is fundamentally a measure of production efficiency—the rate at which inputs are transformed into outputs—and depends critically on how output is defined and measured as well as the cost of the inputs. The Bureau of Labor Statistics (BLS) productivity measures for the health care sector are largely based on service volume measures. As a result, they may understate true productivity growth, since they do not fully capture improvements in health outcomes. Here we present a framework where the output of the sector is the welfare gains from improved health. Other papers in the literature examine the survival aspects of welfare improvement (Dauda et al. (2022) and Cutler et al. (1998)), but this is the first paper to also incorporate the quality-of-life aspects of health improvement by measuring returning to home without an unplanned readmission. Another key contribution of this paper is that we estimate the underlying cost of treatment inputs, while the prior research measures payments, which differ markedly from costs in this regulated environment. To our knowledge, this is the first paper to jointly incorporate a welfare measure of output and cost-based measure of inputs in measuring health care productivity.

While our estimates of productivity gains are large, the magnitude is reasonable in the context of the broader literature. First, acute conditions appear to be an area where productivity gains are especially large, consistent with the findings in Dauda et al. (2022) and Cutler et al. (1998). Productivity gains may be smaller for other conditions. For example, Dunn et al. (2024) examine gains from pharmaceutical innovations across 13 conditions and find distinct results. While they similarly find the quality of treatment improves across all conditions, they find the price of treatment to exceed the benefit for

Medicare population. The statistics are located here: https://www.bea.gov/data/special-topics/health-care.

[7] Moreover, any increase in upcoding is likely more pronounced in MA, where plans receive higher payments based on diagnosis-based risk scores (Geruso and Layton (2020)).

several conditions when the $VQALY is set to $100,000.

Our findings align closely with macroeconomic studies that value improvements in health alongside income growth. Jones and Klenow (2016) consider welfare gains due to both higher incomes and improved health from 1984 to 2006. They find that gains in improved health account for around one third of the welfare gains, even if it accounts for less than 18% of GDP over the period of their study. Other work is also consistent with the large gains in welfare from improved health (Nordhaus (2002), Murphy and Topel (2006), and Hall and Jones (2007)), and other studies suggest that much of the gain is likely attributable to medical care improvements (Cutler et al. (2006) and Cutler et al. (2022)). The study of acute conditions allows us to demonstrate a relatively credible link between medical care inputs and outcomes, supporting the findings in this literature.

The remainder of the paper proceeds as follows. We first describe the methodology, then describe the data, report results, and conclude.

2. Methodology

Our methodological discussion has three parts. First, we discuss the methodology for transforming those risk-adjusted measures into quality-adjusted price indexes and productivity measures. Second, we discuss risk adjustment, as risk-adjusted measures are necessary to attribute costs and outcomes to treatments, and not the health status of patients. Third, we discuss our estimate of long-term outcomes (i.e., life expectancy) and how we connect it to short-term health outcomes (e.g., 90-day post-discharge mortality).

2.1. Quality-Adjusted Price Index and Productivity

Using the risk-adjusted fitted values to measure quality-adjusted life expectancy, $H_{j,t}$, we follow Dauda et al. (2022) to compute a quality-adjusted price $P_{j,t}^{QA}$ for condition j at time t. We apply a Paasche-type price index that applies compensating variation to adjust for quality. Relative to a Laspeyres index, the Paasche index remains bounded to be above zero, even when quality changes are large.[8] The price index measures the expenditure growth required to maintain the same level of utility in period across periods, holding technology constant. To hold technology constant using period t technology, the quality

[8] Dauda et al. (2022) used a Laspeyres, but applied chaining to keep the index between zero and one. Here we apply a Paasche-index and also chaining to reduce large swings in the indexes.

improvements must be added to the treatment price in base period 0.

$$P_{j,t}^{QA} = \frac{S_{j,t}}{S_{j,0} + \$VQALY\Delta H_{j,t}} \tag{1}$$

The components of the index are the risk-adjusted price of disease treatment, $S_{j,t}$, the quality-adjusted life expectancy is $H_{j,t}$, and the dollar value placed on a quality-adjusted life year is $\$VQALY$.[9] Following others in the literature, we apply a range of values for $\$VQALY$, including \$50,000, \$100,000, and \$150,000. Without any change in the quality of treatment, the measure would be a measure of the risk-adjusted price of treatment, $\frac{S_{j,t}}{S_{j,0}}$. Quality improvements lead to reductions in the quality-adjusted price of treatment through the adjustment term in the denominator, $\$VQALY\Delta H_{j,t}$. For example, if $H_{j,t}$ increases by 0.2 and $\$VQALY$ is \$100,000, then the adjustment will be \$20,000.

The quality-adjusted price index is needed for measuring real output and productivity, which we discuss in the following subsection.

2.2. Real Output

Let the number of cases at time t be denoted $N_{j,t}$, so total nominal spending at time t for condition j is $N_{j,t} \cdot S_{j,t}$

We use $P_{j,t}^{QA}$ to adjust nominal spending in period t. Real output in period t is:

$$RealOutput_{j,t} = \frac{S_{j,t} \cdot N_{j,t}}{P_{j,t}^{QA}} = (S_{j,0} + \$VQALY\Delta H_{j,t}) \cdot N_{j,t} \tag{2}$$

The real output in the base period is $S_{j,0} \cdot N_{j,0}$. Therefore, the index of quality-adjusted real output growth $O_{j,t}$ is then:

[9] The exact value to apply for $\$VQALY$ has been estimated in other studies in the literature. See Viscusi (2020) for a recent review of the literature and Kearsley (2024) for recent estimates. Following others in the literature, the risk-adjusted price of treatment $S_{j,t}$ is calculated in 2021 dollars using the PCE aggregate deflator from the BEA.

$$O_{j,t} = \left(\frac{(S_{j,0} + \$VQALY\Delta H_{j,t}) \cdot N_{j,t}}{S_{j,0} \cdot N_{j,0}} \right)$$
$$= \left(1 + \frac{\$VQALY\Delta H_{j,t}}{S_{j,0}} \right) \cdot \frac{N_{j,t}}{N_{j,0}}$$

This is essentially a measure of the growth in the health per case times the growth in the number of cases, so output only increases if health improves or the number of cases increases.

2.3. Cost and Productivity

The second part of the productivity measure is the change in resource inputs, which we measure as the cost, adjusted for inflation, times the number of cases. In this index we account for the change in the input cost of treatment.[10] We apply the BEA aggregate PCE deflator to produce estimates in 2021 dollars. The real risk-adjusted cost per case is $C^R_{j,t}$ so the real cost is $C^R_{j,t} \cdot N_t$ and the growth in the input is $I^R_{j,t}$ and is measured as:

$$I^R_{j,t} = \frac{C^R_{j,t} \cdot N_{j,t}}{C^R_{j,0} \cdot N_{j,0}} \tag{3}$$

The index of productivity is then the ratio of the growth in output to the growth in input: $Productivity_{j,t} = \frac{Y_{j,t}}{I^R_{j,t}}$.

Inserting the formula for growth in real output and growth in real input costs, the number of cases treated cancels out and we are left with:

[10] We apply an economy-wide deflator for the input costs, rather than an input cost deflator, but the practical importance of this difference is small. Our economy-wide deflator grows at a rate of 1.9% annually, while the input cost index from BLS grows at 2.1% annually (Total factor productivity by major industries – December 19, 2025). We chose to use the economy-wide deflator because the price index for medical care inputs is also not adjusted for quality improvements. In this case, medical care price indices would typically show faster price growth, relative to economy-wide inflation, which would tend to increase the productivity measure for the health care sector. For this reason, we apply an economy-wide deflator for costs.

$$Productivity_{j,t} = \frac{(\frac{S_{j,0}+\$VQALY\Delta H_{j,t}}{S_{j,0}})}{C^R_{j,t}/C^R_{j,0}} \tag{4}$$

$$= (\frac{C^R_{j,0} + (C^R_{0t}/S_{j,0})\$VQALY\Delta H_{j,t}}{C^R_{j,t}}) \tag{5}$$

$$\approx (\frac{C^R_{j,0} + \$VQALY\Delta H_{j,t}}{C^R_{j,t}}) \tag{6}$$

where $C^R_{0t}/S_{j,0}$ is the ratio of the cost of treatment in the base period over price of treatment in the base period. The term ratio rescales quality improvements into cost units, allowing us to interpret health gains as the resources required to produce them. If the price is equal to cost $S_{j,0} = C^R_{j,0}$, then the productivity measure is equal to the inverse of the quality-adjusted price index with treatment prices, $S_{j,t}$, values estimated at costs. Evidence from MedPAC suggests that Medicare prices were close to costs at the beginning of our sample in 2002 (Medicare Payment Advisory Commission (MedPAC), 2005). Therefore, we think it is a reasonable proxy to assume overall prices and costs are equal at the beginning of the sample, which we apply in our analysis.

The productivity measure is driven solely by the real cost changes of the input and the quality change of the treatment. If there is no change in quality, then productivity is entirely determined by the cost of the treatment.

Given the deviation of regulated prices from costs, discussed more below, the measure of quality-adjusted costs is another useful metric, which is just the inverse of the productivity measure. If aggregate prices are close to costs, then this may be a reasonable quality-adjusted price index.

$$C^{QA}_{j,t} = \frac{C_{j,t}}{C_{j,0} + \$VQALY\Delta H_{j,t}} \tag{7}$$

It is important to highlight that this method contrasts with the approach applied by BLS across several dimensions, which are outlined in detail at the BLS website. First, the official measure of output for the health care sector is not adjusted for quality. Moreover, it has been argued that a welfare-based adjustment may not fit with BLS concepts of quality adjustment. In particular, Matsumoto et al. (2021) argues that a cost-based adjustment may be preferred to adjust for quality changes and demonstrates how this may be done. The BLS approach also aligns more closely with recent work by Romley et al. (2015) and Romley et al. (2020). Second, we use a measure of cost at the episode level deflated for

the input costs, while the BLS cost estimates are formed from the fundamental inputs (i.e., labor, capital, materials and energy), and a deflator is applied. While there is some difference in how costs are estimated, we view the main empirical difference with that of BLS as primarily coming from the quality-adjusted output measurement.

2.4. Risk Adjustment

We construct risk-adjusted estimates for two key variables: treatment costs and lifetime quality-adjusted life years. We also compute additional risk-adjusted metrics including expenditures, survival for more than 90 days after discharge, and return to home without an unplanned readmission. The steps for risk adjustment are common across the measures. For each outcome, we estimate the following risk-adjustment regression:

$$Y^Z_{i,j,t} = X_{i,j,t} \cdot \beta^Z_j + \gamma^Z_{j,t} + \epsilon^Z_{i,j,t} \tag{8}$$

where $Y^Z_{i,j,t}$ is the relevant dependent variable for patient i, condition j, time t, and outcome variable Z. The $X_{i,j,t}$ are the risk-adjustment variables, $\gamma^Z_{j,t}$ captures the mean risk-adjusted difference (relative to year 0) in the associated outcome measure at time t, and $\epsilon^Z_{i,j,t}$ is the error term. All spending and costs are adjusted for economy-wide inflation using the aggregate PCE price index.

The risk-adjustment term $X_{i,j,t}$ includes a rich set of patient-level controls, including age, sex, race, ethnicity, an income proxy based on residential zip code, and clinically validated risk-adjustment measures described in the data section.

After the equation is estimated the corresponding risk-adjusted index value is predicted as:

$$\hat{Y^Z_{j,t}} = \hat{\bar{X}}_j \cdot \hat{\beta}^Z_j + \hat{\gamma}^Z_{j,t} \tag{9}$$

The fitted value $\hat{X}_{j,t} \cdot \hat{\beta}^Z_j$ uses the coefficient from disease j and measure Z using the entire population sample to construct the average, $\hat{\bar{X}}_j \cdot \hat{\beta}^Z_j$. The fitted coefficient $\hat{\gamma}^Z_{j,t}$ is the fitted value for the change in measure Y for condition j at time t. We refer to $\hat{Y}^Z_{j,t}$ as the risk-adjusted outcome.

For acute conditions, most treatment costs are incurred during the initial hospitalization and 90-day

window that we analyze (see Dauda et al. (2022) for a discussion). However, surviving an acute condition may lead to several additional years of healthy life. Therefore, one of the key variables is the quality-adjusted life expectancy discussed in the next section.

With the above estimates, we derive a variety of risk-adjusted measures from the data. Our key measures for productivity measurement are: risk-adjusted cost ($C_{j,t}$) and risk-adjusted quality-adjusted life expectancy ($QALE_{j,t}$). However, for analytical and comparison purposes, we also compute risk-adjusted spending ($S_{j,t}$), risk-adjusted life-expectancy ($L_{j,t}$), and risk-adjusted return to home without an unplanned readmission ($Home_{j,t}$).

2.5. Long-Term Health Outcomes

The value of medical care for these acute conditions depends on both outcomes associated with the initial hospitalization period as well as long-term health outcomes, conditional on surviving the acute period. This implies an estimate of the long-term survival of the patient is needed. To determine the long-term survival following the 90-day window, Dauda et al. (2022), for example, used the average survival post-episode. For instance, if the average heart-attack patient that survives "short-term" (that is, 90 days beyond discharge) lived for 7 additional years (based on the population sample), then we could add 7 years to those patients surviving the first 90 days.

We take a more precise approach to estimating long-term health outcomes for each patient, leveraging the length and richness of our data. Specifically, we estimate parametric survival models to predict life expectancy conditional on surviving the acute stay. The key idea is to use post-discharge information to translate short-term survival into patient-specific estimates of remaining life expectancy. To do this, we examine an 8-year post-discharge window for each patient-stay to construct a time-to-episode-structured dataset. Here, we record the time from discharge to death for each patient, or whether they survived the 8 years or dropped out of the Medicare sample. Following Kearns et al. (2020), we then fit seven commonly applied parametric survival models for each condition: exponential, Weibull, Gompertz, gamma, log-logistic, log-normal and generalized gamma distributions. We next use the fitted model with the lowest Bayesian information criterion (BIC) for each condition to estimate mean life expectancy by age.

Our parametric survival models include three covariates that are strongly related to both short-term and long-term survival: age, an indicator for any unplanned readmission in the post-discharge period, and an indicator that the patient was discharged home. This allows us to capture, for instance, that saving a younger patient will likely generate more years of additional life than saving an older patient.

While unplanned readmissions and discharge to home have been used as quality-of-life measures in prior studies (Romley et al. (2020)), we use these variables in two ways. They enter directly as quality-of-life adjustments and also help predict long-term survival. We believe this creates a more meaningful long-term health measure and follows in the spirit of the surrogate index methodology (Athey et al. (2019)).

Note that we do not conduct a similar exercise for spending, as we find nearly all of the treatment takes place in the 90-day window, as we discuss further in the descriptive statistics section. However, as a robustness check, we construct a similar lifetime spending effect, assuming post 90-day expenditures are also affected.

2.6. Quality-Adjusted Life Expectancy

We measure health gains by accounting for both additional years of survival and changes in quality of life. Let a quality-adjusted healthy life year be equal to H. The quality-adjusted life expectancy for condition j is measured as $H_{i,j,t} = (90/365) \cdot L_{i,j,t} \cdot Q_{i,j,t} + L_{i,j,t} \cdot R_{i,j} \cdot Q_{i,j,t}$ where $L_{i,j,t}$ is short-term survival (90 days post-discharge) from the episode for condition j, $Q_{i,j,t}$ is the quality-of-life measure for individual i, and $R_{i,j}$ is the predicted number of years alive conditional on surviving condition j for the initial stay and 90-day post-discharge period, discussed in the prior section.[11] This formulation combines short-term survival over the 90-day episode with expected remaining life years, weighted by health-related quality of life. Importantly, $R_{i,j}$ does not change over time, so is not affected by other factors affecting population health or technology over time. The value $Q_{i,j,t}$ is a measure from 0 to 1 where a 1 indicates perfect health and 0 indicates a health state equivalent to death. Following Romley et al. (2020), we assign a weight of $Q = 0.34$ if an individual does not return home or has an unplanned readmission. For all other individuals, we assign a QALY consistent with their age category from Hanmer et al. (2006), where the average QALY for the over 65 population is around 0.8.[12]

[11] We exclude the time period during the initial stay, essentially setting the value of life during the stay to zero. While arguable imperfect, this assumption may be justified by the very low quality of life in an intensive inpatient setting.

[12] Specifically we assign a QALY of 0.826 for those aged 65 to 69, 0.787 for those aged 70 to 79, and 0.753 for those over 80.

3. Data

This research uses a 5% random sample of Medicare beneficiaries from the Centers for Medicare & Medicaid Services (CMS).[13] Our sample includes patients aged 65 or older who are continuously enrolled in traditional fee-for-service Medicare (both Parts A and B) during the episode. We focus on fee-for-service enrollees and Medicare Part A and B claims, including (but not limited to) CMS's Medicare Provider Analysis and Review File (MedPAR). The MedPAR file aggregates all claims from an inpatient stay in a single line.

As mentioned previously, we focus on nine medical episodes that are defined as starting with an inpatient stay: (1) acute myocardial infarction (heart attack), (2) congestive heart failure, (3) pneumonia, (4) gastrointestinal hemorrhage (GI bleed), (5) hip fracture, (6) stroke, (7) lower-extremity joint replacement (LEJR), (8) chronic obstructive pulmonary disease (COPD), and (9) coronary artery bypass graft surgery (CABG). These episodes are identified by applying established algorithms to diagnosis codes and other information (e.g., major diagnostic category) reported in the MedPAR record.

For each of the conditions studied, the episode of care begins with admission to an acute-care short term hospital, and the end of the episode is marked after the 90-day post-discharge period or until death. Death dates are available from CMS's Beneficiary Summary Files, in which the beneficiary death date has been validated through Social Security Administration data.

Similar to Romley et al. (2020), we use inpatient data to identify unplanned hospital readmissions by applying the algorithm developed for CMS and described in for Outcomes Research & Evaluation (2014). Additional details of this algorithm and how it is applied can be seen in the appendix and in Romley et al. (2020). Similar to Romley et al. (2020), the quality metric also includes information on whether an individual is discharged to home during the patient's last facility-based claim, which is closely related to another quality metric adopted by CMS.[14] While remaining in an institutionalized setting or having a readmission to the hospital is not equivalent to mortality, it does have an effect on the quality of life. To capture this difference in quality, we adopt the approach of Romley et al. (2020) and measure quality in a quality-adjusted life year (QALY) scale from 0 to 1, where 1 indicates a year of life in perfect health and 0 is death. The prior work of Romley et al. (2020) uses external data and methods from Cutler et al. (1997) to determine a QALY decrement. Following that work, the QALY decrement that is applied is 0.66, indicating the QALY is 0.34 for those experiencing an unplanned readmission or for

[13] One important limitation of our work, relative to Romley et al. (2020), is that we have only a 5% sample, while they worked with a 20% sample. While our sample is smaller, this does not appear to affect the key results, as we obtain qualitatively very similar estimates when applying the same methodology to our data, as shown in the appendix.

[14] This quality metric was adopted by CMS under the Improving Medicare Post-Acute Care Transformation Act of 2014.

those not discharged to the community during the episode window. Additional details and discussion are included in the appendix and in Romley et al. (2020).

The claims data do not directly provide information on costs.[15] As in prior work (Romley et al. (2020)), we estimate costs of all stays (including *any* readmission to hospital or inpatient post-acute care, such as skilled nursing, after the initial discharge) by multiplying a measure of utilization—specifically, charges—by cost-to-charge ratios for each provider-year.[16] The latter are obtained from financial reports provided by the different institutional providers that are required to submit financial information to CMS, which are often referred to as cost reports in the literature. Importantly, the charges are not used in our estimation procedure, we just use the cost-to-charge ratio to assign costs to utilization. Cost reports for hospitals, skilled nursing facilities and inpatient rehabilitation facilities are linked to patient episodes based on the patient's scrambled ID and the provider's Medicare Provider Number.

Ideally, we would estimate real costs using the actual inputs for each episode (e.g., labor, materials, and capital services). However, this detailed information is not available at this level of granularity. Instead, the cost reports provide a feasible method of transforming information on utilization into resource costs at the hospital level. There are several advantages to using cost reports to estimating costs. They are mandatory for all facilities that participate in Medicare, the reports are reviewed and sometimes audited by CMS, and the cost reports follow a standardized measurement. The cost reports are extensively used by MedPac, an independent congressional agency that advises the U.S. Congress, to understand the costs and finances of hospitals and other facilities, as well as the Congressional Budget Office (e.g., Hayford et al. (2016)). In addition, cost reports have been used extensively in prior academic studies including recent work to measure input costs and financial outcomes (Gaynor et al. (2026) and Gupta et al. (2024)) as well as past related work (Romley et al. (2015) and Romley et al. (2020).

We also include the cost of physician care, home health care, and durable medical equipment, following Romley et al. (2020). For these services, we do have cost-to-charge ratios, so we use total payments to proxy for costs. We do not include prescription drugs, because the Part D program launched during our analysis window (in 2006).[17] Nominal costs are converted to real dollars based on an economy-wide PCE price deflator, as discussed previously, which is close in value to the BLS deflator used for input costs in the health care sector.[18] Additional details regarding the cost estimates are available in Romley

[15] The claims also report payment, including the allowed amount, which is the total reimbursement the hospital and physicians receive for the services provided.

[16] Where necessary, cost-to-charge ratios and other indices are calendarized on a days-weighted basis across provider reporting periods.

[17] Moreover, much of the drug costs are fixed development costs, so it is not clear that the payments observed on prescription claims would accurately reflect the economic costs of these treatments, as discussed in Dunn et al. (2024).

[18] In our methodology, we assume that the price-cost margin for physician services is zero in 2002, then use the inflation component of the annual Medicare Economic Index to quantify real growth in physician costs.

et al. (2020).

All outcomes are risk-adjusted using information from claims and enrollment files. In particular, these data contain information on the patient's age, sex, race, ethnicity, and detailed information on health conditions of each patient based on disease diagnosis (and procedure) codes reported on the initial stay. The claims data also report the zip code in which the patient resides, which we link to zip code-level information from the U.S. Census on several socioeconomic characteristics that proxy for patient severity, include the poverty rate, income, share of elderly institutionalized, among other variables.

Diagnosis codes are aggregated into the Charlson comorbidity index, which has been shown to predict death within 12 months (Charlson et al., 1987). In addition, we apply publicly available and clinically validated risk measures for six of the nine episode types.[19] For each of these episodes, we use a patient-specific probability of death during the initial stay, produced as part of Inpatient Quality Indicators developed by the Agency for Healthcare Research and Quality (AHRQ).

Prior to October 2015, the disease codes were based on the global ICD-9 taxonomy, but in late 2015 the disease classification was updated to the ICD-10 scheme.[20] While we are able to estimate productivity and its components for all eight conditions pre-2015, we are only able to estimate productivity for six conditions post-2015 based on availability of ICD-10 risk-adjustment models from AHRQ. Even though many of the conditions are defined similarly pre- and post-2015 (such as pneumonia, heart failure, and heart attack), because of the change in diagnosis codes, we treat both the spending allocation and risk adjustment as distinct across the two periods. The ICD-10 changes were substantial enough that the risk adjustment and health outcomes are not directly comparable across the 2015 threshold.

4. Descriptive Statistics

Overall, the data we compile include around 1.9 million episodes over the 20-year period from 2002 to 2021. Some descriptive statistics for the population are shown in Table 1. The average age of the population in the data is around 80 years old, with a higher share female (62%) and white (88%). Around 87% of individuals survived the hospitalization and 90-day post-discharge period, but of those that survive, only about 75% return home without a readmission. One notable feature of the descriptive statistics is that the total cost per episode exceeds the payment by about 8%.

[19] We are missing risk adjustment for CABG, hip and knee replacement, and COPD

[20] There are several differences between the ICD-9 and ICD-10 disease classification systems, but perhaps the most notable difference is the level of detail. The ICD-9 has around 13,000 codes, while the ICD-10 has around 68,000 codes. For purposes of risk adjustment in this study, both systems contain sufficient detail for commonly applied risk adjustment.

It is worth highlighting that we focus on the 90-day period for spending, as the 90-day spending appears to capture all the costs associated with the episode. To demonstrate this, Figure A15 in the appendix shows average costs in the months before, during, and after the episode. We find the main increase in costs above trend occurs over the 90-day period, similar to the findings in Dauda et al. (2022). We explore the sensitivity of this assumption in the robustness section of the paper.

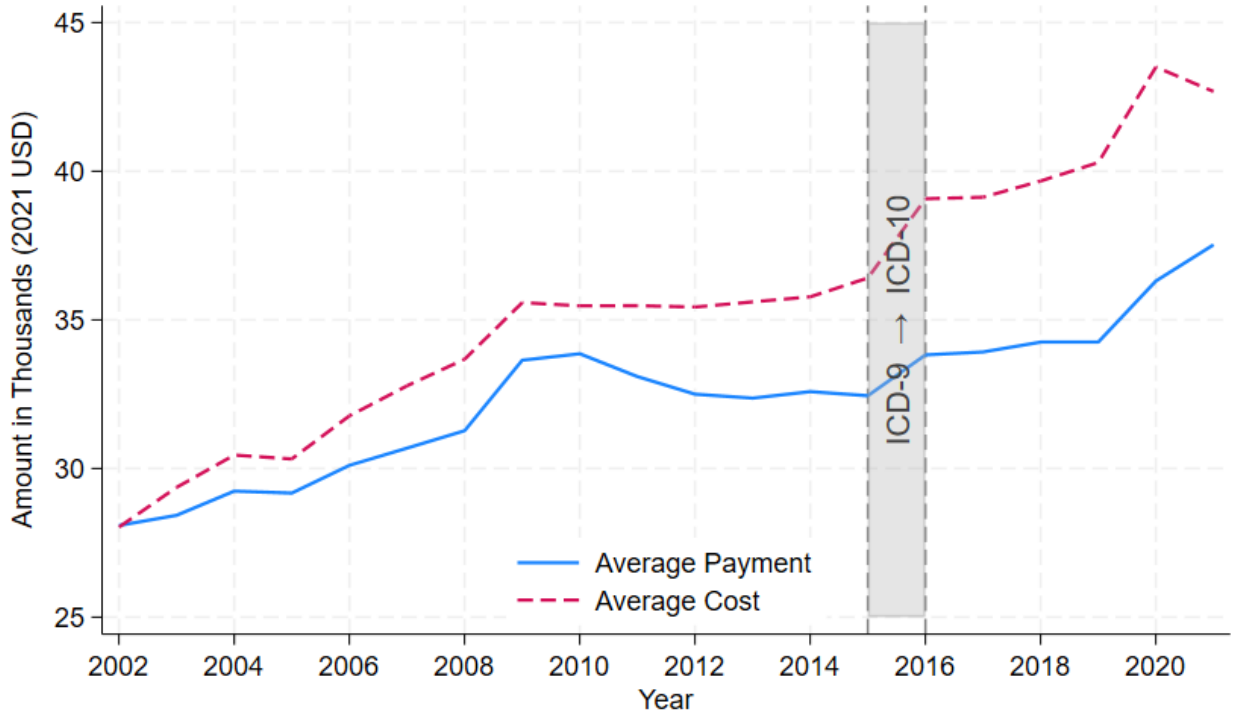
Table 1. Descriptive Statistics

	Mean	Median	SD	P10	P90	*N*
Age	79.25	79.00	8.23	68.20	90.43	1,916,575
Female	0.62	1.00	0.49	0.00	1.00	1,916,575
Race: White	0.88	1.00	0.33	0.00	1.00	1,916,575
Race: Black	0.08	0.00	0.28	0.00	0.00	1,916,575
Race: Hispanic	0.02	0.00	0.12	0.00	0.00	1,916,575
Income in thousands (zip code)	43.56	39.85	16.34	27.40	65.10	1,894,597
Share less than HS educ. (zip code)	0.19	0.18	0.11	0.07	0.34	1,894,502
Share elderly instit. (zip code)	0.06	0.05	0.06	0.00	0.12	1,894,255
Share phys. disabled (zip code)	0.29	0.29	0.08	0.20	0.38	1,894,189
Share mentally disabled (zip code)	0.11	0.10	0.05	0.06	0.17	1,894,189
Risk adjustment score	0.03	0.01	0.06	0.00	0.08	1,300,929
Charlson comorbidity index	2.09	2.00	1.89	0.00	5.00	1,916,575
Payment per episode (2021 thous. $)	31.75	24.34	26.42	10.11	61.87	1,916,575
Cost per episode (2021 thous. $)	34.40	26.07	31.48	8.75	68.81	1,916,575
Alive at discharge	0.87	1.00	0.34	0.00	1.00	1,916,575
Discharged home (alive)	0.75	1.00	0.43	0.00	1.00	1,916,575
30-day unplanned readmission	0.14	0.00	0.35	0.00	1.00	1,916,575
Alive 90-days post-discharge, discharged home, no readmission	0.67	1.00	0.47	0.00	1.00	1,916,575

Notes: This table shows basic descriptive statistics based on variables in the data. Payment information is based on the allowed amount taken directly from the claims data, while cost estimates combine claims data with information from cost reports and other sources to estimate costs. Age, female, and race variables are taken directly from the claims data. Income in thousands is based on the median household income of the patient's zip code of residence, matching on Zip Code Tabulation Areas from the Census Bureau. Other zip code level variables include share elderly, share physically disabled, and share mentally disabled. The risk adjustment score and Charlson comorbidity index are computed based on the algorithms described in the text and prior work by Romley et al. (2020). Both the total payment amount and total cost are reported in 2021 dollars using the aggregate PCE deflator from the BEA.

Figure 1 reports the trend in average payment and average cost based on Medicare claims data and cost information, relative to economy-wide inflation. The vertical lines in 2015 and 2016 indicate the transition in disease classification from 2015 to 2016, indicating that this jump is likely related to the disease classification change. While both the costs and payments rise relative to economy-wide inflation, the costs rise substantially faster than payments. This finding is consistent with external studies that show similar trends, and demonstrate that focusing on Medicare payments would overstate productivity

Figure 1. Trends in Average Payments and Average Cost in 2021 Dollars, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



Notes: The figure shows the average estimated cost and average spending per patient using the claims for our nine conditions studied in this paper. Estimates are deflated by the aggregate PCE deflator so are relative to economy-wide deflation. The figure shows that the gap widens over the sample period. The ratio is 1 in the beginning of the sample, while payments are 12–15% below cost by the end of the sample. The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. For this reason the estimated change from 2015–16 is not necessarily economically meaningful. Costs and payments are adjusted to 2021 dollars using the aggregate PCE price index.

growth.[21]

In 2019, the costs for this bundle of episodes is 15% above the payments. This gap between payments and costs highlights the need for using costs to measure productivity gains for the health sector contrasting with prior work that focused solely on Medicare payments, likely overstating productivity growth (Dauda et al. (2022)).

[21] MedPac reports indicate negative profit margins of over 10 percent for Medicare patients in recent years (https://www.medpac.gov/wp-content/uploads/2025/03/Mar25_Ch3_MedPAC_Report_To_Congress_SEC.pdf), while prior MedPac reports in 2003 showed margins close to zero. Consistent with this divergence between costs and payments, prices from private payers have grown considerably faster than Medicare rates over this period, as shown by estimates from the BLS PPI program.

4.1. Risk-Adjusted Descriptive Statistics

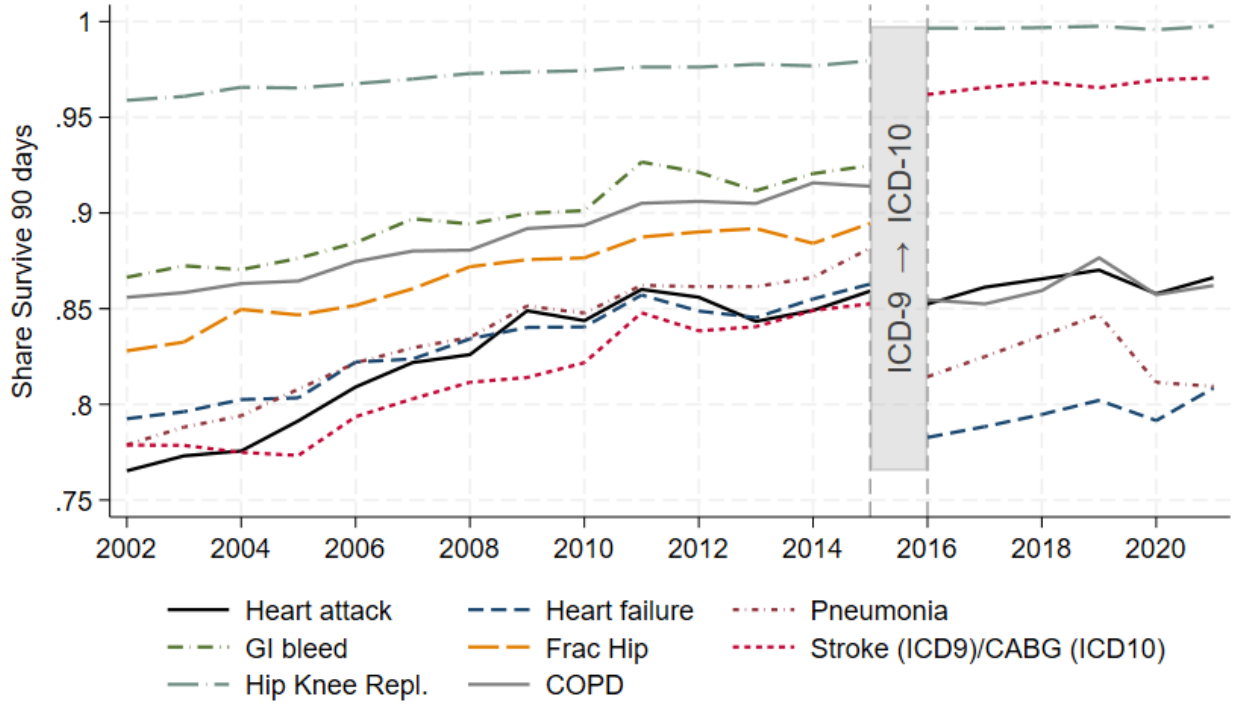
An important step in forming the productivity statistics is to first risk-adjust all of the key outcome variables, including survival, return-to-home without a readmission, payment, cost, and life time QALYs. Due to the change from ICD-9 to ICD-10 in 2015, we estimate all of the outcome variables over two periods, first from 2002 to 2015, and then from 2016 to 2021. Over each period, the risk adjustment is estimated flexibly by condition. Each condition includes a third-order polynomial for age, a dummy if the patient is female, an interaction of female and age, income quartile, and race and ethnicity dummies, in addition to socioeconomic characteristics of the patient's zip code. Charlson comorbidity index and AHRQ risk adjustment variables are also included. In addition, we interact income quartile with age, race, ethnicity, Charlson comorbidity index, and AHRQ risk adjustment with the understanding that treatment and risk adjustment may be different across socioeconomic groups. In a robustness section we show that our key findings are robust to the risk-adjustment methodology, specifically we run our estimates without controlling for any diagnostic risk adjustments.

The risk-adjusted survival trends are shown in Figure 2. The left panel shows risk-adjusted survival pre-2015, which shows large increases for many conditions with a range of 5–10 percentage points for most conditions. The gains in knee replacement were only about 2%, but that is the only condition where the key outcome variable is arguably return-to-home. The gains for the six conditions shown in the right panel for the 2016–2019 period are positive for all the conditions, but the rate of improvement is not as large. Interestingly, while risk-adjusted life expectancy increased over the 2016–2019 period, population life expectancy stayed relatively flat at around 78 years from 2016 to 19.[22] These estimates are consistent with the theory that the decline in life expectancy trends may deviate from medical care outcomes due to external population health factors.

For most conditions there is a sharp decline in survival from 2019 to 2020 with the onset of the COVID pandemic in 2020 for four of the six ICD-10 conditions, but we observe some rebound in survival from 2020 to 2021 for three out of the four conditions showing a decline. The notable exception is pneumonia, which is the condition mostly clearly tied to complications from COVID. COVID complicates the analysis because the risk-adjusted survival before COVID is not comparable to risk-adjusted survival in 2020 with the introduction of a new condition. COVID introduces some uncertainty over the productivity of the health sector over this time horizon. One approach would be to treat 2019–2020 as a structural break. However, outside of the COVID shock itself, the populations observed in 2016–2019 and 2020–2021 are broadly comparable. Based on this assumption, the survival generally grew over this time horizon. For our analysis below, we assume that these risk-adjusted survival estimates are accurate, including

[22] Vital Statistics Rapid Release, Number 023 (August 2022)

Figure 2. Risk-Adjusted Survival, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



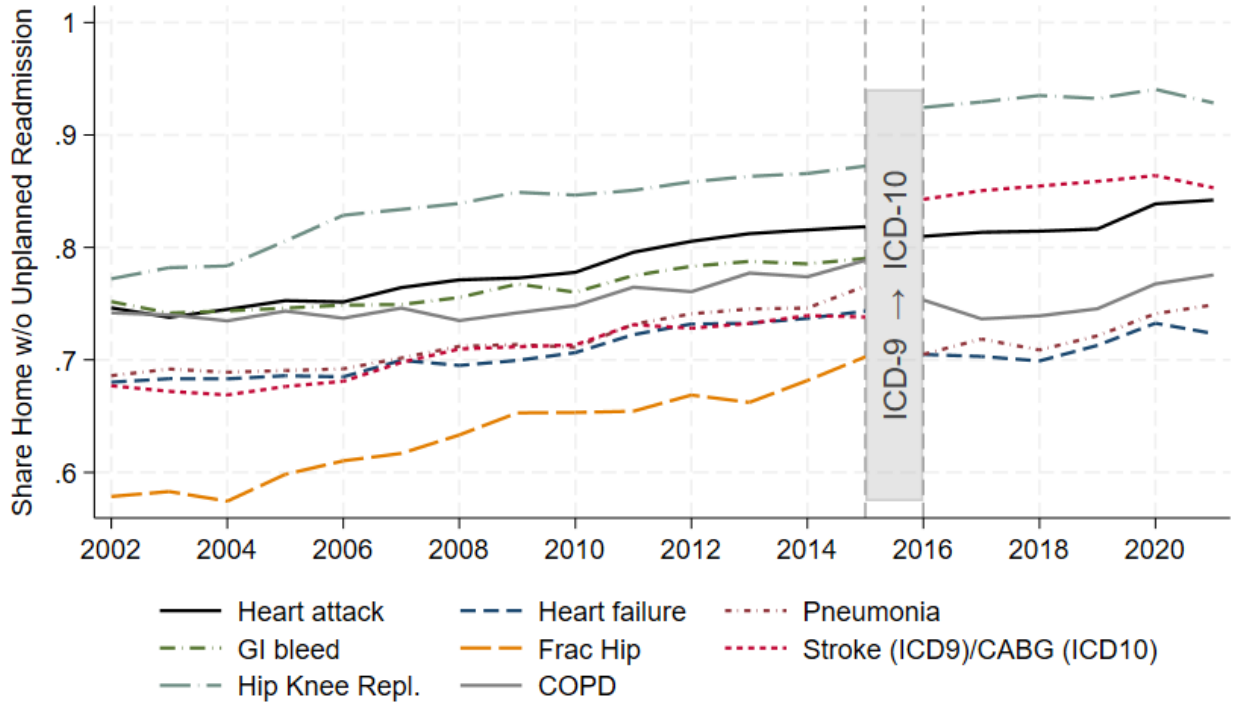
Notes: The figure shows the risk-adjusted survival of patients pre-2015 (left half of figure) and post-2015 (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. For this reason the estimated change from 2015–2016 is not economically meaningful. The risk-adjustment described in greater detail in the text controls for a variety of variables including age, sex, predicted income, race, and risk measures based on the patient's conditions.

the 2019–2020 period, and view these estimates as likely understating the health improvements due to medical care.

We next examine the estimates on return to home without an unplanned readmission. The left panel of Figure 3 shows an upward trend in return to home without an unplanned readmission for all conditions pre-2015. This increase was particularly large for hip fractures. In the post-2015 ICD-10 period, the trend was relatively more flat, with some reduction in return-to-home without unplanned readmission from 2020 to 2021 for some conditions, which could also potentially be affected by the introduction of COVID in 2020. The overall trend highlights the importance of accounting for this aspect of the patients quality-of-life. Both of the components of this index, return to home and unplanned readmissions, show improvement over this period as reported in Figures A5 and A4 of the appendix.

Figure 4 show the risk-adjusted costs for the conditions over this time period. There are a few points to

Figure 3. Risk-Adjusted At Home Without Unplanned Readmission, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



Notes: The figure shows the risk-adjusted return to home without readmission for patients that survive the episode. The figure shows estimates for patients pre-2015 (left half of figure) and post-2015 (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. For this reason, the estimated change from 2015 to 2016 is not economically meaningful. The risk adjustment, described in greater detail in the text, controls for a variety of variables including age, sex, predicted income, race, and risk measures based on the patient's conditions.

note. First, the costs for all conditions rose during the ICD-9 pre-period. The costs have been relatively flat for these conditions since around 2 009. The cost growth estimates are adjusted for economy-wide inflation, so this flattening signifies cost trends matching economy-wide inflation trends.[23]

Risk-adjusted survival and returning to home without an unplanned readmission are transformed into expected lifetime QALYs as described previously in section 2.6. An important part of this transformation is predicting future survival of each patient. Figure A14 in the appendix depicts the relationship between these three covariates and conditional life expectancy for acute myocardial infarction and pneumonia. Life expectancy is uniformly higher for patients discharged home without unplanned readmissions, and it is uniformly lower for patients not discharged home with an unplanned readmission. While the direction of these effects is not surprising, the magnitudes are quite large. For example, for a heart attack patient, a 65-year-old that returns to home without an unplanned readmission is expected to live about 13 years, while a 65-year-old that has a readmission and does not return to home is predicted to live about 7.5 years.

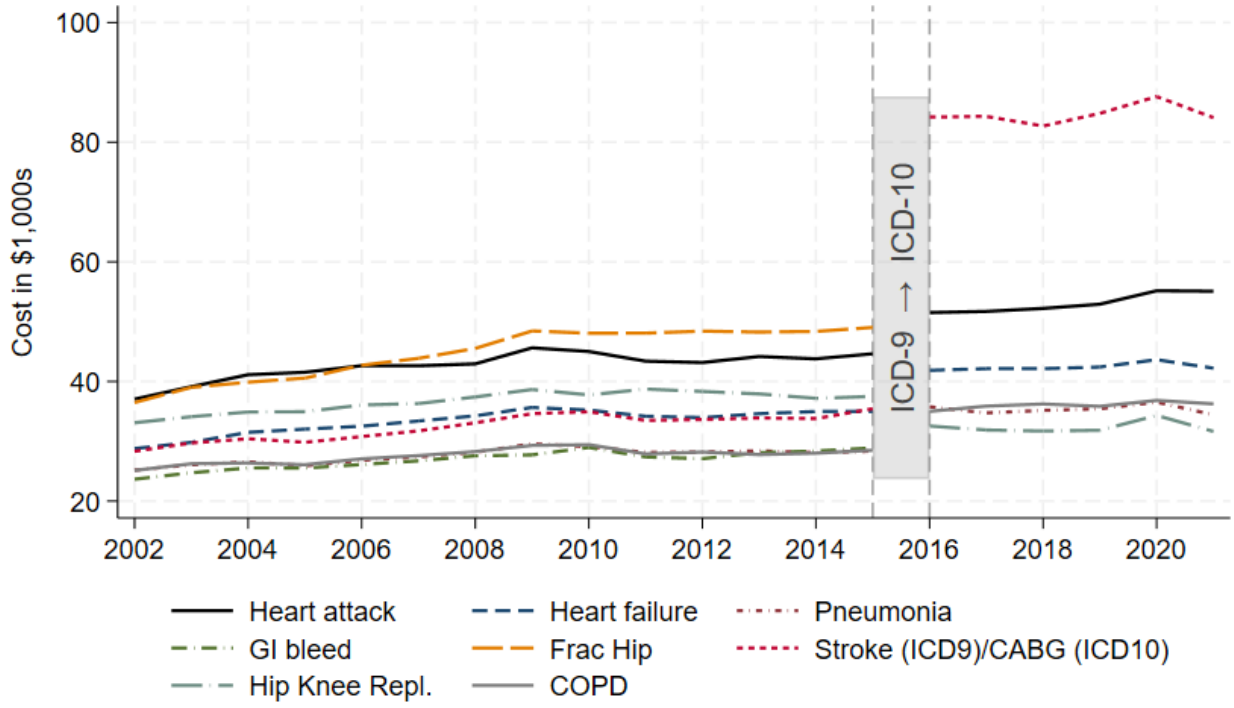
5. Results

This section presents the main productivity results. We present the quality-adjusted cost and its inverse, our measure of productivity. Figure 5 shows the quality-adjusted cost for all of the conditions using a $VQALY of $100,000. Quality-adjusted costs decline sharply for all conditions in the pre-2015 period. This is consistent with the quality-adjusted prices reported in Dauda et al. (2022) for three health conditions heart attack, heart failure, and pneumonia. This paper expands those findings to five other conditions, and incorporating gains in quality-of-life by reflecting a greater share of patients returning to home without an unplanned readmission. We also use costs, rather than payments used in prior work, which is an important distinction given that regulated Medicare prices seem to deviate and grow more slowly than costs. Interestingly, one of the sharpest declines in quality-adjusted price over the period are for hip and knee replacement, where return to home without an unplanned readmission is a key outcome, rather than mortality. The post-2015 period presents new results to the literature by showing large quality-adjusted price declines over the 2016 to 2019 period.

While there is some increase in the quality-adjusted cost from 2019 to 2020 for pneumonia, this change should be interpreted cautiously, as it was likely greatly impacted by COVID. Despite the influence of COVID having a negative impact on outcomes over 2019–2020, arguably biasing the cost indexes upward and productivity downward, we still find the quality-adjusted cost index declining for five of the six conditions tracked over this period.

[23] The corresponding estimates for risk adjusted payments are shown Figure A3 in the appendix

Figure 4. Risk-Adjusted Cost, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



Notes: The figure shows the risk-adjusted costs for each condition studied. The figure shows estimates for patients pre-2015 (left half of figure) and post-2015 (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. For this reason, the estimated change from 2015 to 2016 is not economically meaningful. The risk adjustment, described in greater detail in the text, controls for a variety of variables including age, sex, predicted income, race, and risk measures based on the patient's conditions. Estimates are reported in 2021 dollars applying the aggregate PCE price index.

The productivity estimates show the inverse of the cost index measure in Figure 6. The productivity gains are quite large for many of the conditions, reflecting the gains in health, relative to the growth in the cost of the episodes. The productivity for heart failure rises the least, but the productivity gain was still substantial with a gain of over 100% over this period. Productivity gains were largest for pneumonia and hip and knee replacement, with productivity gains of over 400% over this time period. The change from ICD-9 to ICD-10 may affect this interpretation as condition definitions shift, but we generally see large productivity gains in the pre-period from 2002 to 2015. We also observe productivity gains in the post-period 2016–2019 prior to COVID.

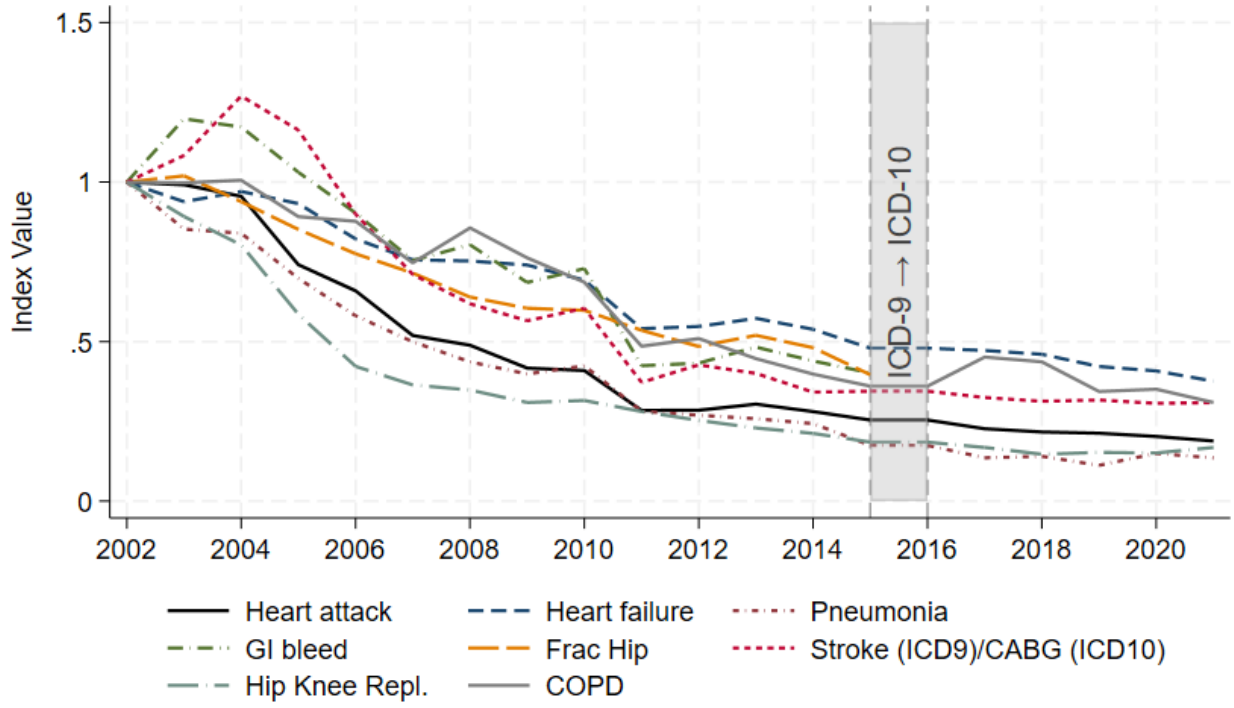
The productivity decline for hip and knee replacement is somewhat surprising over the 2019–2021 period. While the QALY and costs estimates are fairly flat, there is a slight reduction in QALY outcomes in 2021. However, this may be due to selection effects around the pandemic, as there was a dramatic shift away from total joint replacements in the inpatient setting toward the outpatient setting (Piple et al. (2023)). While this shift started in 2018 when CMS moved total joint replacement off the inpatient-only list, it accelerated around the pandemic.

To examine the overall productivity over the 2002 to 2021 period, we apply chaining to the Paasche formula, and we combine estimates across conditions by weighting each condition based on its expenditure share in the prior period. We explore the aggregate estimates based on our central estimate of a \$VQALY of \$100,000, but also \$50,000 and \$150,000. The results of the estimate are shown in Figure 7. The top panel shows the decline in the quality-adjusted cost index, while the bottom panel shows gains in productivity. The cost decline for the central estimate is about 7% per year, corresponding to a productivity gain of 7.5% a year. The results are highly sensitive to the value placed on a statistical life year, showing a productivity gain of 11.0% per year if the \$VQALY is \$150,000 and about 3.5% per year if the \$VQALY is \$50,000. The productivity growth slows a bit post 2015, but the cause of this is uncertain, as the change in the studied conditions related to the ICD-9/ICD-10 change may cause some of this shift, and the 2020 pandemic also disrupts these estimates. Overall, the finding of large gains in productivity growth are robust to several alternative specifications discussed in the next section.

The magnitudes of these estimates are large, but it is important to consider them in the context of the broader health care sector. We are focusing on conditions where the productivity gains may be especially large. Recent work by Dunn et al. (2024) studies 13 health conditions and associated innovations for those conditions. Using a VSLY of \$100,000, they find that for six of the 13 conditions, the price actually exceed the benefits, suggesting that one should be cautious in applying the findings for the nine conditions studied here to the broader sector.[24] Moreover, the methodology for measuring productivity

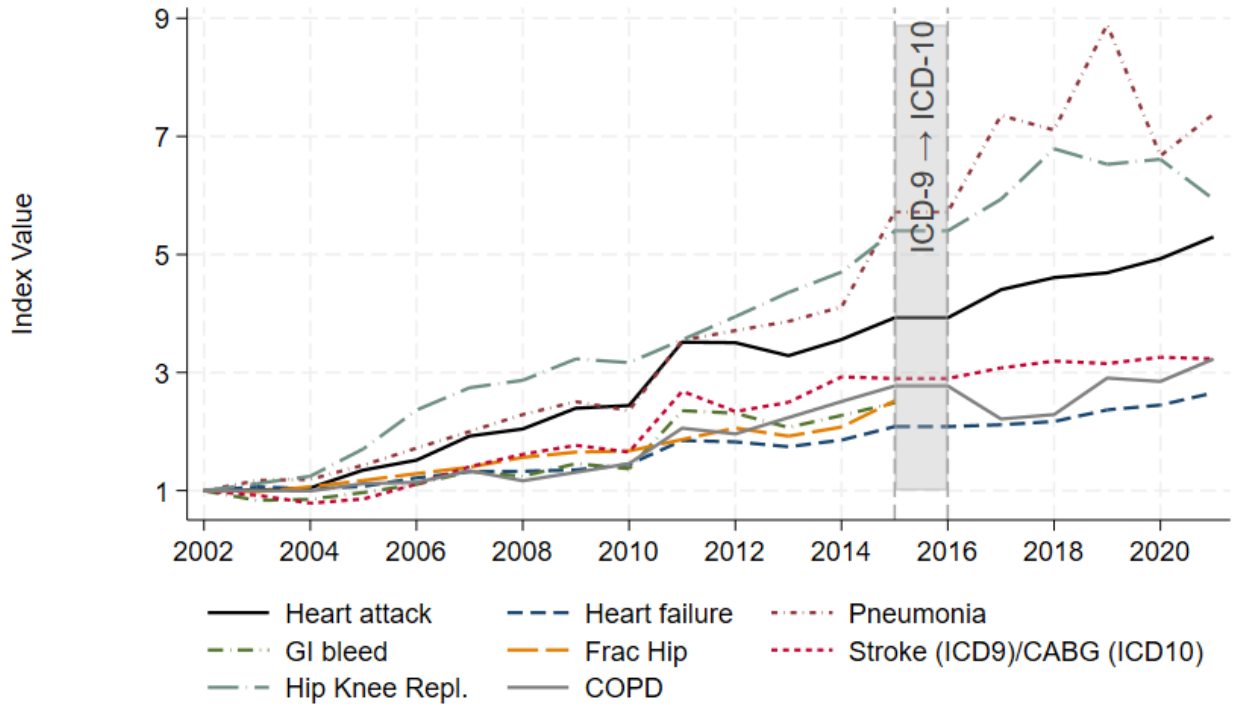
[24] Although they also find benefits exceed the costs, much of the benefit goes toward producer surplus for many of the innovations that they study, so that productivity rises, even if consumer surplus falls.

Figure 5. Quality-Adjusted Cost Index Adjusted With Change in Lifetime QALY ($VQALY 100,000), 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



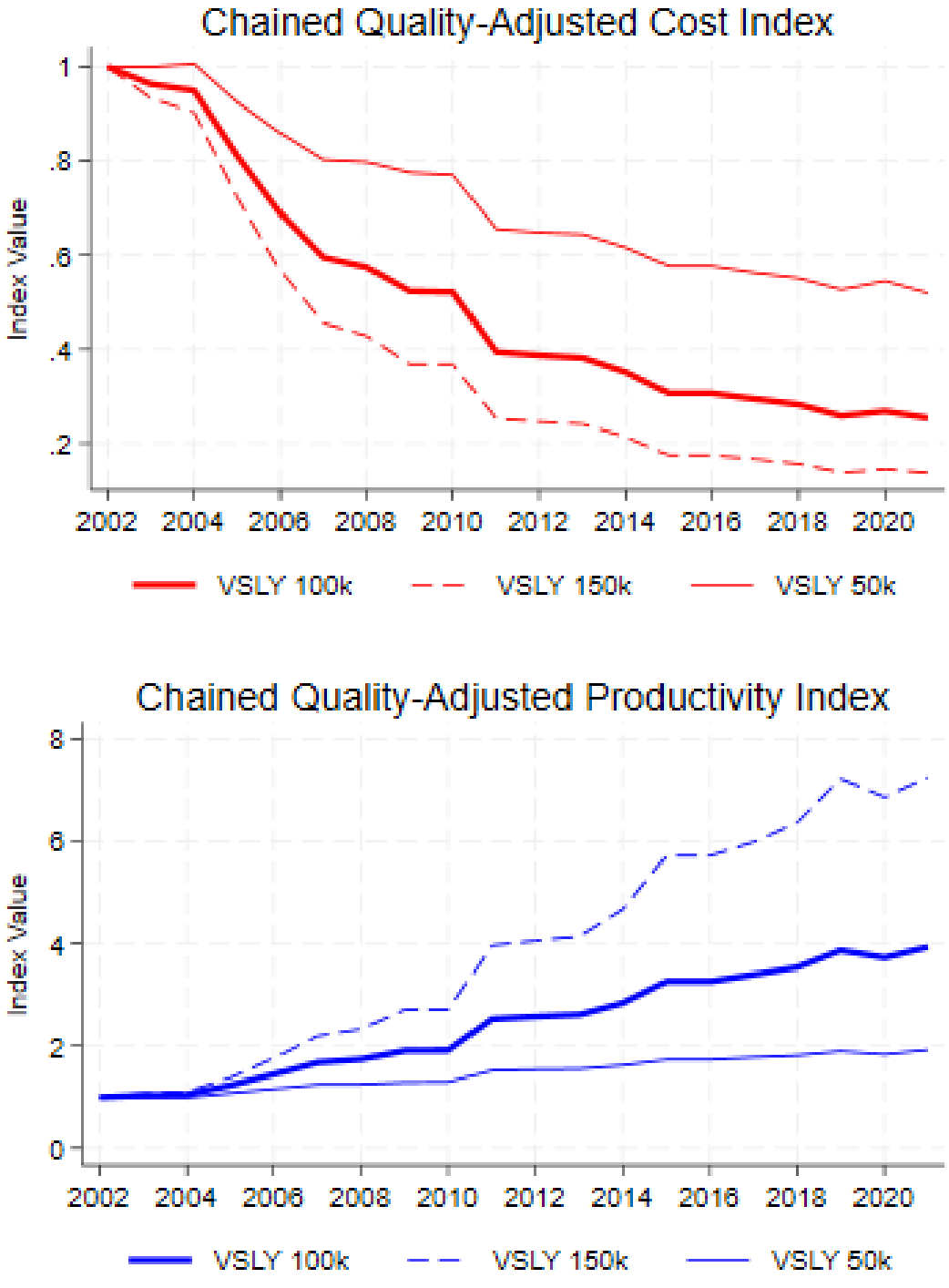
Notes: This figure shows chained index applying the risk-adjusted cost index. The figure shows the pre-2015 indexes (left half of figure) and post-2015 indexes (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. For this reason the estimated change from 2015-16 should be interpreted with caution. The index is held at no change from 2015 to 2016 for those conditions observed in pre- and post- period. The dollar amounts are adjusted for economy-wide inflation using the aggregate PCE deflator.

Figure 6. Quality-Adjusted Productivity Index Adjusted With Change in Lifetime QALY ($VQALY 100,000), 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



Notes: This figure shows chained risk-adjusted productivity index. The figure shows the pre-2015 indexes (left half of figure) and post-2015 indexes (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. For this reason the estimated change from 2015-16 should be interpreted with caution. The index is held at no change from 2015 to 2016 for those conditions observed in pre- and post- period. The dollar amounts are adjusted for economy-wide inflation using the aggregate PCE price index.

Figure 7. Aggregate Chained Quality-Adjusted Cost and Productivity, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



Notes: Chained measures applying $100,000 VSLY, $50,000 VSLY, and $150,000. The top panel shows the quality-adjusted cost, and the bottom panel shows the productivity measure. Index chaining is weighted by the expenditure share of each condition in the prior period. Note that different conditions are used in the chaining in the pre-2015 period and the post-2015 period due to the ICD-9 and ICD-10 change. The index is held at no change from 2015 to 2016 for those conditions observed in pre- and post- period.

also matters, which likely explains why our results differ from BLS productivity estimates as well as Matsumoto et al. (2021) and Romley et al. (2020) that apply adjust for quality based on the cost of treatment. As discussed in Sheiner and Cutler (2024), a quality adjustment based on the cost of producing the quality change may understate the total welfare gain, for example, due to technological constraints. To apply this idea in our paper, the risk-adjusted cost of a heart attack patient was $56k in 2021, while the risk-adjusted QALY of the patient was 5.7 QALYs, so the average cost per QALY was about $1k (=$56k/5.7QALY). The average cost of the QALY is extremely low in this case, but Sheiner and Cutler (2024) argues that additional health is not purchased at this price because it is not possible due to technological constraints. Sheiner and Cutler (2024) argues that in the case where technology is constrained, it is better to use the marginal value to consumers, rather than the cost.[25] As another example, some technologies, such as an aspirin for a heart attack patient, may lead to large health improvements and welfare gains at no cost, which we argue should be viewed as a major productivity improvement.

Using a cost-based approach, Romley et al. (2020, 2015) finds productivity gains, but they are substantially smaller than those found in this paper. Similarly, Matsumoto et al. (2021) generates a cost-based quality adjustment by relating quality measures, including measures on acute health outcomes and also satisfaction scores, to the cost of treatment, carefully accounting for hospital selection. They find that accounting for quality would lead to lower quality-adjusted price indexes, but the magnitude of the adjustment is substantially smaller than the welfare-based adjustment presented here.

To determine the broader implications of our estimates for the aggregate health sector productivity, we compute the share of spending for our nine selected conditions using estimates from the BEA health care economic statistics. These statistics are useful for this purpose as they allocate spending for the health care sector into 260 condition categories. There are seven health conditions that roughly correspond to the condition categories studied in this paper.[26] The conditions we study account for 3.8% of overall spending in 2021. Assuming no productivity growth in official health-sector measures,[27] we multiply the spending share of our selected conditions by our estimated annual productivity gains. This implies an aggregate productivity bias of about –0.29 percentage points per year.

[25] This may be an extreme case, as the marginal costs per QALY may be increasing, so the average cost may not be an accurate reflection of the marginal cost of the health gain, as noted in Dauda et al. (2022). However, even when considering the marginal cost, the marginal benefits may not be equal to the marginal cost due to market distortions.

[26] The conditions in the BEA health care economic statistics are partitioned into 260 Clinical Classification Software (CCS) categories, a classification system constructed by AHRQ. The CCS categories that most closely correspond to the conditions in this paper include congestive heart failure (108), acute myocardial infarction (100), fracture of neck of femur (hip) (226), osteoarthritis (203), gastrointestinal hemorrhage (153), acute cerebrovascular disease (109), and pneumonia (122).

[27] The BLS annual growth in total factor productivity for the health care and social assistance sector was 0.05% per year from 2002 to 2021 suggesting approximately no productivity gain over this period based on official measures.

6. Robustness Checks

We conduct a number of robustness checks, shown in the appendix and described here. First, in Figure A6, we evaluate the quality-adjusted price index based on Medicare payments, rather than cost, and we find quality-adjusted prices fall a bit more quickly, consistent with costs increasing more rapidly than Medicare payment trends. Second, in Figure A7, we apply the Laspeyres type index chained, and we get results similar to our main findings and similar to Dauda et al. (2022) who apply this index. Third, rather than using a survival model to predict mortality for each patient after the 90-day period (e.g., survival of 74 year old without an unplanned readmission), we use the average survival for those that survived the initial 90-day period. The results, shown in Figure A8, are similar to the main findings, but with quality-adjusted prices growing slightly faster. For this robustness check, we also compute the aggregate price index and productivity, shown in Figure A10, which shows productivity growth to be a similar magnitude, but a bit slower than our main estimates.

As another robustness check we examine how potential effects of spending outside the 90-day window may affect our estimates. As shown in Figure A15 we find strong evidence that most cost occur within the 90-day window, and costs fall below trend post 90 days. However, here we apply an alternative assumption to investigate the sensitivity of our results.[28] To obtain our alternative costs estimates we use information on average monthly costs before, during, and after the episode period, shown in Table A5. In particular, rather than considering the upward trend in cost pre- episode shown in Figure A15, we assume that the difference in cost from the pre- to the post- period is related to the episode. We then annualize this additional cost (column 5 of Table A5), we then construct a ratio of annual costs to the 90-day costs around the episode that we use to predict lifetime costs.[29] The results of this change on the quality-adjusted cost indexes are shown in Figure A9, and the aggregate affects are shown in Figure A11. The productivity estimates remain large, but the magnitude of the increase is smaller than in the baseline, with a 6.2% annual growth in productivity.

For the last two robustness checks, we adjust the controls that are applied in the specification. While we view risk adjustment as essential for accurate measurement, one concern with applying risk adjustment is that the diagnosis codes used in the adjustment may be affected by changes in how practitioners code (e.g., upcoding), so that the diagnosis codes may not accurately reflect the illness of the patient. We are not aware that this is an issue for the conditions that we study or, more importantly, that

[28] We view this sensitivity analysis as particularly useful, as adding some additional costs to may be justified, as at costs not reflected in these claims data, such as increases in at home care provided by family members (Wolff et al. (2025)).

[29] For instance, for heart attacks this suggests that annual post 90 day spending is 16% of the 90-day episode spending. To obtain post-90 day lifetime costs we calculate: (90-day cost individual i)*(Share Annual Post-90 day spending)*(Life Expectancy for Individual i).

this practice changes over time. However, to check the robustness of the findings, we only control for factors unaffected by diagnostic practices, such as age, sex, race, and socioeconomic variables from the zip code data, such as income. The aggregate estimates, shown in Figure A12, are substantially noisier but still show large productivity gains over the sample period of about 4% a year until 2019. The lack of risk adjustment has a substantial effect over the 2020 and 2021 period, as controlling for diagnosis over the COVID period may be especially important. Productivity falls considerably over the 2020 and 2021 period, but still remains higher than the 2002 level. Finally, as another robustness check, we include additional controls. Specifically, we include characteristics of the hospital and the demographics of the surrounding area for the hospital where the patient is admitted, similar to the controls in Romley et al. (2019). The aggregate estimates are shown in Figure A13. The estimates change relatively little compared to the main specification.

We explored a number of other robustness checks that showed similar results, although not included in the appendix.[30]

7. Conclusion

We quantify sustained improvements in medical-care productivity for nine conditions over 2002–2021, combining risk-adjusted outcomes with a utility-based framework that values gains in both longevity and quality of life. By measuring resource use with costs (rather than payments) and linking short-run outcomes to long-run QALYs, we find large declines in quality-adjusted prices and corresponding growth in productivity—on the order of about 7.5% per year using our central assumptions. These results extend prior work by covering more conditions and a longer horizon (including the ICD-10 period), and they remain qualitatively robust across a number of assumptions, including changes in the value of QALYs. While the productivity gain is large, these conditions account for 3.8% of overall spending, suggesting an aggregate productivity bias of –0.29 percentage points in official reporting.

There are a few notable limitations to our analysis. First, the ICD-9/ICD-10 transition creates breaks in comparability. We attempt to address this issue with chaining at the aggregate level, as the differences in coding across thresholds seem to be too large to be reconciled. Second, COVID-19 introduced a nontrivial shock to case mix and mortality risk. While our approach attempts to mitigate this issue through risk adjustment, it seems that COVID-19 likely has a measurable effect on productivity, especially

[30] For example, we apply the same price index formula as our main specification, but we do not apply chaining and find prices fall only slightly faster than our main results. We've run the specification not accounting for quality of life, and only examining survival, and we obtain qualitatively similar findings in the aggregate.

for pneumonia. For this reason, our estimates likely understate changes in the productivity of the health care sector over this period.

There are several promising areas for future research. First, our research primarily focuses on aggregate measures, but there may be important heterogeneity in trends across the population. Future research should assess differential productivity effects across populations. Second, this paper focuses on services received over an episode where payments are typically paid on a FFS basis. This is not the case for bundled payment programs that pay for a fixed bundle per episode and aim to reshape the incentives of providers to improve outcomes and costs (Agarwal et al. (2020)). While these papers often look at distinct performance metrics (e.g., costs or readmissions), an interesting avenue of work would be to see how these programs affect productivity.

References

3M™ Health Information Systems (2026). 3M™ APR DRG Classification System and 3M™ APR DRG Software. https://multimedia.3m.com/mws/media/4784150/3m-apr-drg-fact-sheet.

Agarwal, R., J. M. Liao, A. Gupta, and A. S. Navathe (2020). The impact of bundled payment on health care spending, utilization, and quality: A systematic review: A systematic review of the impact on spending, utilization, and quality outcomes from three centers for medicare and medicaid services bundled payment programs. *Health Affairs 39*(1), 50–57.

Agency for Healthcare Research and Quality (2026). AHRQ IQI Technical Documentation, Version v2025. https://qualityindicators.ahrq.gov/measures/IQI_TechSpec.

Athey, S., R. Chetty, G. W. Imbens, and H. Kang (2019). The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely. Technical report, National Bureau of Economic Research.

Charlson, M. E., D. Carrozzino, J. Guidi, and C. Patierno (2022, January). Charlson comorbidity index: A critical review of clinimetric properties. *Psychother. Psychosom. 91*(1), 8–35.

Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases 40*(5), 373–383.

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

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

Cutler, D. M., E. Richardson, T. E. Keeler, and D. Staiger (1997). Measuring the health of the us population. *Brookings papers on economic activity. Microeconomics 1997*, 217–282.

Cutler, D. M., A. B. Rosen, and S. Vijan (2006). The value of medical spending in the united states, 1960–2000. *New England journal of medicine 355*(9), 920–927.

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

Dunn, A., L. Fernando, and E. Liebman (2024). How much are medical innovations worth? a detailed analysis using cost-effectiveness studies. Technical report, Bureau of Economic Analysis.

for Outcomes Research & Evaluation, Y. N. H. H. S. C. C. (2014). 2014 measure updates and specifications report hospital-wide all-cause unplanned readmission: version 3.0.

Gaynor, M., A. Sacarny, R. Sadun, C. Syverson, and S. Venkatesh (2026). The anatomy of a hospital system merger: the patient did not respond well to treatment. *Review of Economics and Statistics 108*(1), 272–281.

Geruso, M. and T. Layton (2020). Upcoding: evidence from medicare on squishy risk adjustment. *Journal of Political Economy 128*(3), 984–1026.

Grieco, P. L. and R. C. McDevitt (2017). Productivity and quality in health care: Evidence from the dialysis industry. *The Review of Economic Studies 84*(3), 1071–1105.

Gupta, A., A. La Forgia, and A. Sacarny (2024). Turbocharging profits? contract gaming and revenue allocation in healthcare. Technical report, National Bureau of Economic Research.

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

Hanmer, J., W. F. Lawrence, J. P. Anderson, R. M. Kaplan, and D. G. Fryback (2006). Report of nationally representative values for the noninstitutionalized us adult population for 7 health-related quality-of-life scores. *Medical Decision Making 26*(4), 391–400.

Hartman, M., A. B. Martin, D. Lassman, A. Catlin, and N. H. E. A. Team (2026). National health care spending increased 7.2 percent in 2024 as utilization remained elevated: Article examines national health care spending in 2024. *Health Affairs*, 10–1377.

Hayford, T., L. Nelson, and A. Diorio (2016). *Projecting Hospitals' Profit Margins Under Several Illustrative Scenarios*. Congressional Budget Office.

Jackson, C. (2016). flexsurv: a platform for parametric survival modeling in r. *Journal of statistical software 70*, 1–33.

Jones, C. I. and P. J. Klenow (2016). Beyond gdp? welfare across countries and time. *American Economic Review 106*(9), 2426–2457.

Kearns, B., J. Stevens, S. Ren, and A. Brennan (2020). How uncertain is the survival extrapolation? a study of the impact of different parametric survival models on extrapolated uncertainty about hazard functions, lifetime mean survival and cost effectiveness. *Pharmacoeconomics 38*(2), 193–204.

Kearsley, A. (2024). Hhs standard values for regulatory analysis, 2024. Technical report, Assistant Secretary for Planning and Evaluation, Health and Human Services.

Matsumoto, B. et al. (2021). *Producing Quality Adjusted Hospital Price Indexes*. US Department of Labor, US Bureau of Labor Statistics, Office of Prices and ....

Medicare Payment Advisory Commission (MedPAC) (2005). Hospital inpatient and outpatient services. In *Report to the Congress: Medicare Payment Policy*, pp. 45–78.

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

Nordhaus, W. D. (2002). The health of nations: the contribution of improved health to living standards.

Piple, A. S., J. C. Wang, G. J. Bouz, B. C. Chung, C. K. Mayfield, M. K. Richardson, D. A. Oakes, J. R. Lieberman, A. B. Christ, and N. D. Heckmann (2023). The persistent effects of the covid-19 pandemic on total joint arthroplasty changes in practice patterns in the united states from 2020 to 2021. *The Journal of Arthroplasty 38*(8), 1438–1443.

Quan, H., V. Sundararajan, P. Halfon, A. Fong, B. Burnand, J.-C. Luthi, L. D. Saunders, C. A. Beck, T. E. Feasby, and W. A. Ghali (2005, November). Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med. Care 43*(11), 1130–1139.

Romley, J. A., A. Dunn, D. Goldman, and N. Sood (2019). Quantifying productivity growth in health care using insurance claims and administrative data. *Big Data for 21st Century Economic Statistics*.

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

Romley, J. A., D. P. Goldman, and N. Sood (2015). Us hospitals experienced substantial productivity growth during 2002–11. *Health Affairs 34*(3), 511–518.

Sheiner, L. and D. M. Cutler (2024). Measuring prices and productivity in the health care sector. In *The Measure of Economies: Measuring Productivity in an Age of Technological Change*, pp. 225–262. University of Chicago Press.

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

Syverson, C. (2011). What determines productivity? *Journal of Economic literature 49*(2), 326–365.

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

Wolff, J. L., J. C. Cornman, and V. A. Freedman (2025). The number of family caregivers helping older us adults increased from 18 million to 24 million, 2011–22: Article examines growth in the number of family caregivets providing help to older adults in the us. *Health Affairs 44*(2), 187–195.

Online Appendix:
Quantifying Productivity Growth in Medical Care

Appendix A Data Construction and Risk Adjustment

This paper builds on the episode construction and data processing framework developed in earlier work by Romley et al. (2020). While the main text summarizes the core empirical strategy, this appendix briefly describes the construction of episodes of care, the use of Medicare cost reports to measure treatment costs, the identification of readmissions, and the risk-adjustment variables. Readers seeking additional technical details should consult the earlier paper.

A.1 Episode Construction

Episodes of care are constructed using Medicare fee-for-service claims. Each episode begins with an admission to a short-term acute-care hospital for a condition identified using validated diagnosis or procedure codes. The conditions examined include acute myocardial infarction, heart failure, pneumonia, gastrointestinal bleeding, hip fracture, stroke, lower-extremity joint replacement, chronic obstructive pulmonary disease, and coronary artery bypass graft surgery (CABG).

Consistent with earlier work, an episode starts with the index hospitalization and continues for 90 days after discharge or until death, whichever occurs first. This episode window aligns with CMS bundled-payment models that use a similar 90-day post-discharge horizon. Death dates are obtained from the Medicare Beneficiary Summary File, which links administrative records with validated mortality data from the Social Security Administration.

Only beneficiaries aged 65 or older who are continuously enrolled in traditional fee-for-service Medicare (Parts A and B) during the episode are included in the analysis.

A.2 Identification of Readmissions

Hospital readmissions are identified using the claims-based algorithm developed by CMS and for Outcomes Research & Evaluation (2014). This algorithm classifies readmissions as planned or unplanned and excludes scheduled follow-up procedures. Unplanned readmissions occurring within the relevant episode window are incorporated into the episode and contribute to total episode costs and outcome measures.

Episodes may also include services delivered after discharge from the initial hospitalization, including additional inpatient stays, outpatient services, physician visits, skilled nursing facility care, home health services, hospice care, and durable medical equipment.

A.3 Measuring Treatment Costs Using Cost Reports

Medicare claims data report utilization measures such as hospital charges but do not directly measure the underlying cost of care. Following Romley et al. (2020), we convert utilization measures into costs using financial information reported in Medicare cost reports submitted annually by participating providers.

Institutional providers—including hospitals, skilled nursing facilities, and inpatient rehabilitation facilities—report cost-to-charge ratios in their cost reports. We link these provider-level ratios to claims using the Medicare provider identifier and compute episode-level costs by multiplying reported charges by the corresponding cost-to-charge ratio. This approach is widely used in the literature to approximate resource costs from administrative claims data.

Physician services are measured using Relative Value Units (RVUs), which quantify the resources required to provide specific medical services. RVUs are converted into costs using the Medicare physician conversion factor and adjusted for geographic differences.

Total episode costs therefore include all institutional and professional services occurring within the episode window. Nominal costs are converted to real terms using the PCE economy-wide inflation measure from BEA.

A.4 IQI-Based Risk Adjustment

For seven of the nine conditions analyzed in this paper, we incorporate risk adjustment based on AHRQ's Inpatient Quality Indicators (IQIs).(Agency for Healthcare Research and Quality (2026)) AHRQ supported the creation of these measures for the purpose of evaluating hospital quality using administrative discharge data. AHRQ tasked teams of clinical experts with developing condition-specific risk-adjustment models and made the resulting algorithms, including risk coefficients, publicly available.

The IQI models produce predicted mortality probabilities based on patient-level characteristics observed in standard inpatient records. For each episode type with an available IQI model, we compute the predicted probability of survival for each hospitalization using AHRQ's IQI software developed. These models incorporate detailed clinical information including diagnosis codes, procedure codes, patient demographics and severity classifications.

A key input into the IQI risk-adjustment models is the All Patient Refined Diagnosis Related Group (APR-DRG) classification system, particularly 3M's APR-DRG risk-of-mortality (ROM) scale.[1](3M™ Health Information Systems (2026)) The APR-DRG grouper aggregates diagnosis and procedure codes into clinically meaningful categories and assigns patients to severity levels that capture differences in expected mortality risk. Although the APR-DRG grouping algorithm is proprietary, a limited-use version is distributed through the AHRQ IQI software and can be applied directly to administrative claims data.

Following Romley et al. (2020), we apply IQI version 6.0.2 for use with ICD-9 coding, followed by a version updated for ICD-10 coding (specifically, the current version v2024). The predicted probability of survival generated by the IQI model is then aggregated to the hospital-year level by averaging predicted survival across all episodes treated at a given hospital in a given year. This variable serves as a measure of the underlying clinical severity of patients treated at that hospital.

Including the IQI predicted survival measure allows us to control for systematic differences in patient health status that may vary across hospitals and over time. In particular, if hospitals begin treating sicker patients over time, failing to adjust for this change would bias estimates of productivity downward. The IQI risk adjustment helps ensure that changes in measured outcomes reflect changes in treatment effectiveness rather than changes in patient severity.

For hip and knee replacement and COPD, there is no IQI model of mortality risk. However, these

[1] 3M™ asserts that this system is "the most widely used SOI and ROM adjustment tool today and the standard for adjusting large volumes of data to account for differences related to the individual's true condition. As a result, organizations can focus on the differences in clinical care and generate apples-to-apples comparisons of quality."

conditions are included in CMS's unplanned readmission algorithm, and so we analyze them, using the algorithm to identify these hospitalizations.

For all conditions, we also apply alternative measures of patient health. These include the widely used Charlson–Deyo comorbidity index, developed for the purpose of predicting 12-month mortality, applied to the diagnoses on the initial hospital claim.(Quan et al. (2005), Charlson et al. (2022)) We further include demographic and Census-based geographic area characteristics summarized in Table 1. These additional risk metrics are particularly important for the 2 conditions for which an IQI mortality risk model is not available.

Appendix B Quality-Adjusted Life Expectancy Additional Detail

We construct survival analysis datasets for each acute condition using Medicare beneficiary enrollment files linked to inpatient claims. The estimation sample consists of patients admitted from 2002 to 2012, allowing for a minimum 8-year follow-up window extending through 2021. For each patient stay, we record the time from hospital discharge to death (measured in years), with observations censored if the patient survived the full follow-up period. Survival times are discretized to quarter-year intervals, with times at or below zero set to 0.125 years. The sample is restricted to patients who survived the 90-day post-discharge acute period, as we seek to estimate long-term survival conditional on surviving the initial recovery window.

We estimate parametric survival models using the flexsurv package in R, which provides a unified framework for fitting and comparing multiple parametric distributions (Jackson (2016)). For each condition, we fit seven commonly used survival distributions—exponential, Weibull, Gompertz, gamma, log-logistic, log-normal, and generalized gamma—via maximum likelihood estimation. All models include three covariates: patient age at admission, an indicator for any unplanned readmission within 30 days of discharge, and an indicator for discharge to home (versus skilled nursing facility or other institutional setting). Model selection is based on the Bayesian Information Criterion (BIC), with the lowest-BIC model selected for each condition.

Using the selected parametric model, we compute mean life expectancy over a 30-year time horizon for a grid of covariate values spanning ages 65 to 95 and all combinations of the home-discharge and readmission indicators. Point estimates are obtained from the fitted survival function. This approach

yields condition- and age-specific life expectancy estimates that account for heterogeneity in post-acute outcomes, enabling us to assign patient-specific expected remaining life years based on observable discharge characteristics.

Appendix C Additional Descriptive Statistics

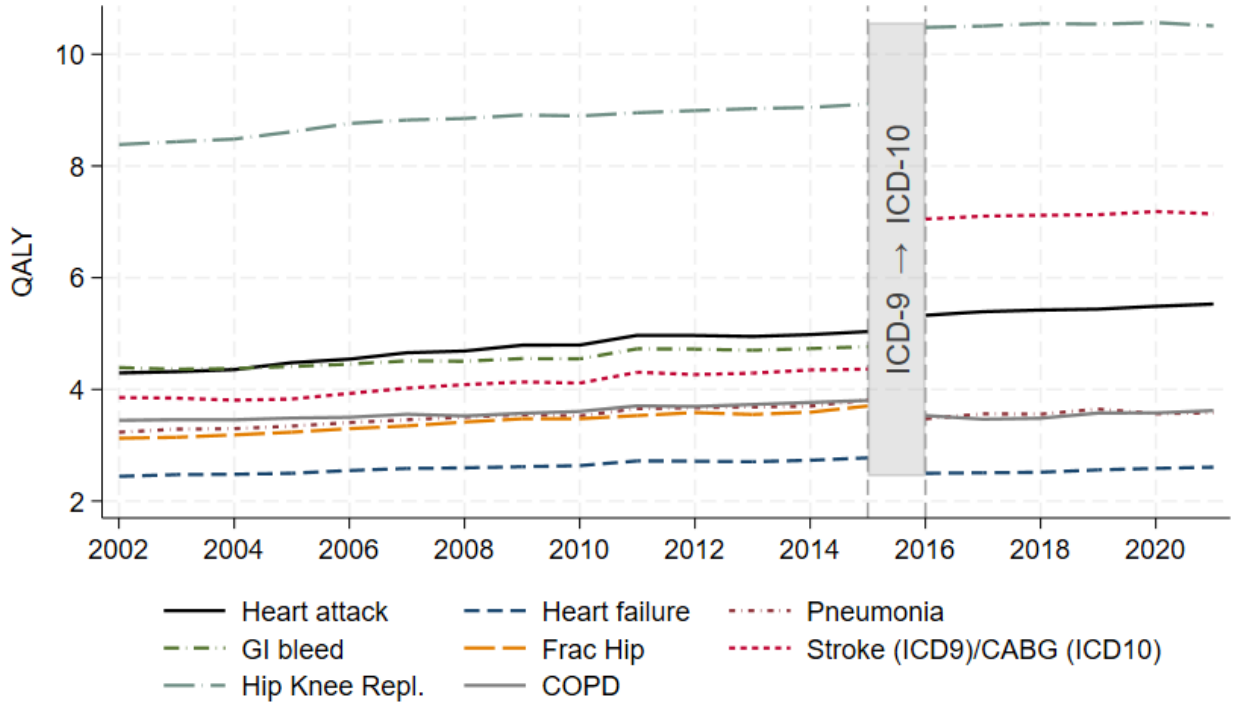
This section shows a few additional descriptive statistics. Table A1 shows a count of the number of episodes for selected years of the sample. The table shows the sample size is relatively stable until the switch from ICD-9/ICD-10 and then COVID, where sample sizes shrank. There are a few contributing factors related to the sample size. First, the ICD-9/ICD-10 switch affected directly the definitions of the conditions. Second, the ongoing switch from Medicare FFS to Medicare Advantage reduced sample size throughout the entire period (FFS claims are needed to measure outcomes and costs during the 90-day post-discharge window). Third, in 2021, we did not observe the subsequent year, so observations near the end of the year must be dropped. While the sample size falls, we have numerous demographic and condition specific controls for risk factors, which we think helps standardize the population, even as the sample size shifts.

Table A1. Descriptive Statistics Bundle Count, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)

	2002	2005	2010	2014	2016	2019	2020	2021
Heart attack	11,316	9,670	8,281	7,653	7,954	7,414	6,226	4,305
Heart failure	27,143	24,638	19,293	16,783	18,005	21,207	16,187	12,482
Pneumonia	27,297	25,988	17,036	14,378	13,099	13,788	11,456	8,358
GI bleed	12,826	12,161	9,597	8,679	—	—	—	—
Hip fracture	8,886	9,035	8,199	7,604	—	—	—	—
Stroke	7,710	9,238	8,211	7,790	—	—	—	—
CABG	—	—	—	—	2,182	2,261	1,808	1,320
Hip/knee replacement	18,935	21,957	21,677	21,437	16,321	16,184	8,140	4,564
COPD	16,393	15,966	14,238	11,051	12,291	10,351	5,818	3,374

Notes: This table shows a count of the number of conditions by year for selected years. The change in the number of episodes from 2014 to 2016 is due to the ICD-9 to ICD-10 switch. AHRQ discontinued production of its risk adjustment models for GI bleed, hip fracture and stroke during the ICD-10 period. The change in the number of episodes across the sample and in particular in 2021 is due to several factors. First, since we do not observe the subsequent year of data in 2022, the observations in the last quarter are dropped, since we do not have a complete 90-day window. Second, COVID may have led to fewer incidence of the various conditions. Third, there has been a trend in switching from FFS to Medicare Advantage, reducing the sample size in our data. Finally, preventative drugs such as statins for high cholesterol and antihypertensive drugs have likely led to fewer episodes.

Figure A1. Risk-Adjusted Lifetime QALY Trends by Condition, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



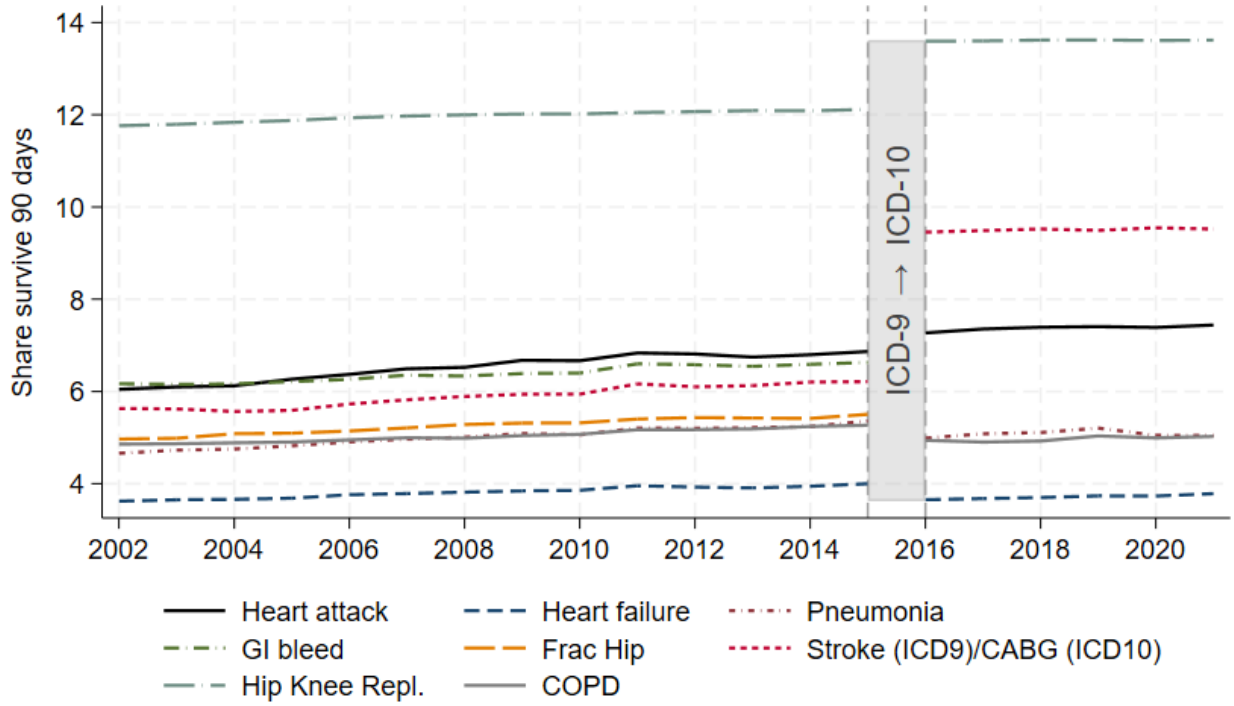
Notes: The figure shows the average estimated risk adjusted lifetime QALYs by condition. The figure shows the trends pre-2015 (left half of figure) and post-2015 (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. For this reason the estimated change from 2015–2016 is not economically meaningful.

Figure A1 shows the trends for the lifetime QALY across conditions. As one might expect, lifetime QALYs for hip and knee replacement is substantially larger than the other conditions, as it is typically a nonfatal condition. Also note that the QALYs tend to rise across all conditions.

Figure A2 is the risk-adjusted life expectancy across all conditions. As can be seen, the life expectancy is higher than the lifetime QALYs, as it is not downweighted for limited quality of life.

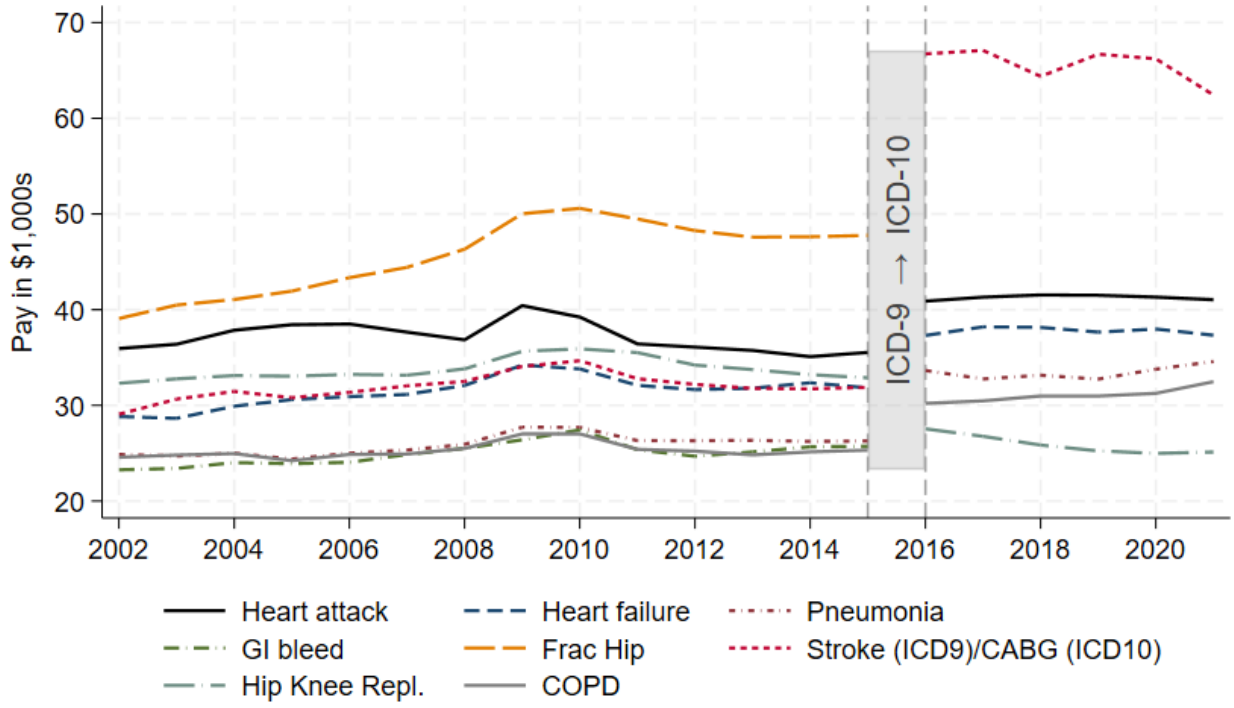
Figure A3 shows risk-adjusted average payment by medical condition. The relative trends in payments tend to be close to the trends in cost reported in Figure 4 of the main paper.

Figure A2. Risk-Adjusted Years of Life Trends by Condition, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



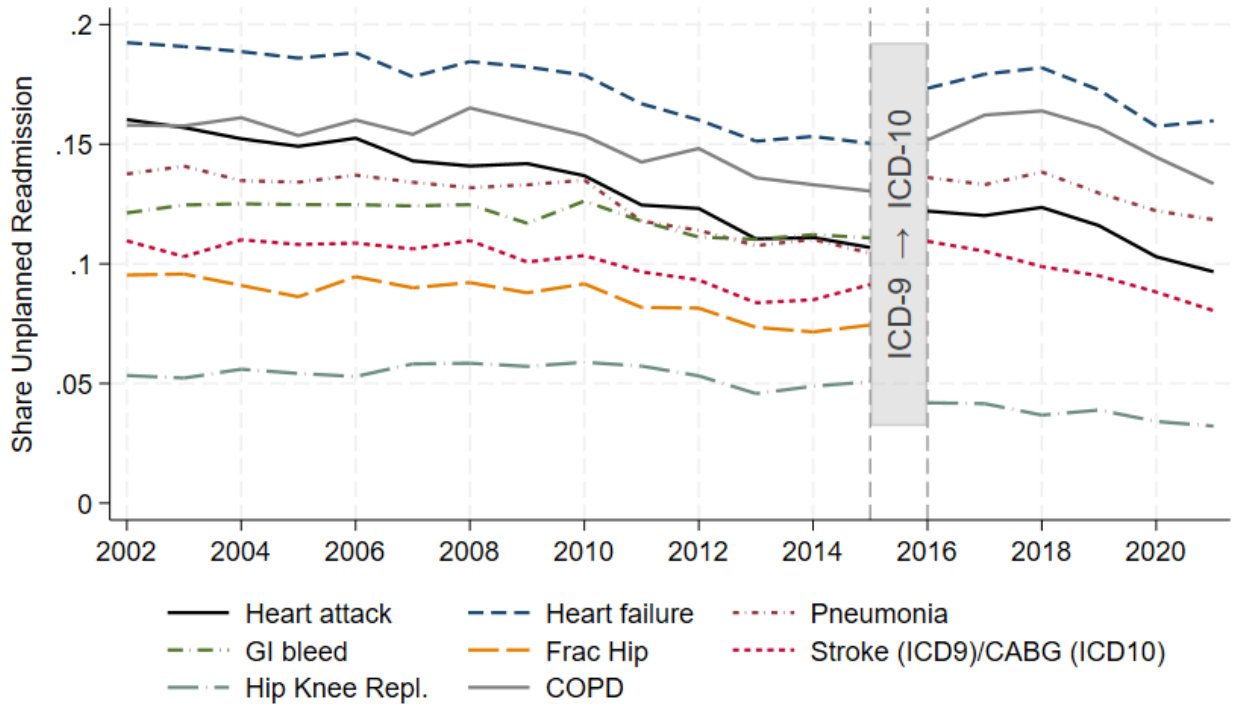
Notes: The figure shows the average estimated risk adjusted lifetime years of life by condition. The figure shows the trends pre-2015 (left half of figure) and post-2015 (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. For this reason the estimated change from 2015–2016 is not economically meaningful.

Figure A3. Risk-Adjusted Payment by Condition, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



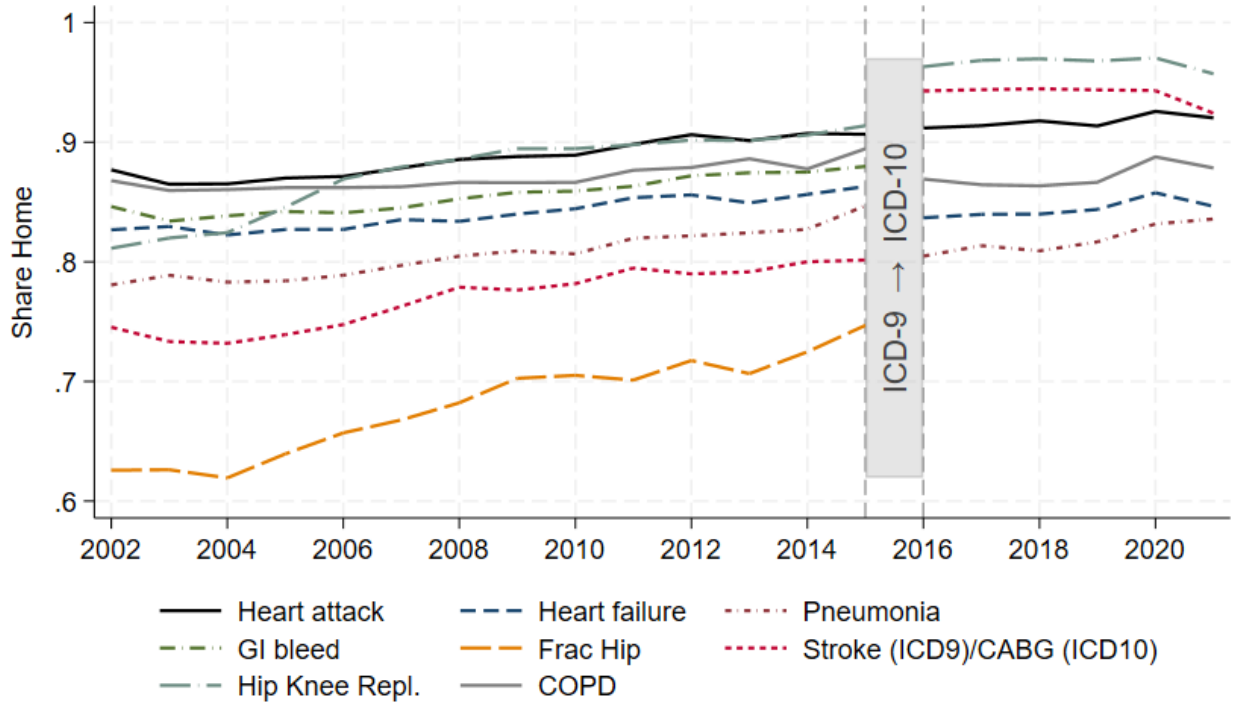
Notes: The figure shows the average estimated risk-adjusted 90-day spending based on Medicare payments. The figure shows the trends pre-2015 (left half of figure) and post-2015 (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. For this reason, the estimated change from 2015–2016 is not economically meaningful. The dollar amounts are adjusted for economy-wide inflation using the aggregate PCE price index.

Figure A4. Risk-Adjusted Unplanned Readmission by Condition, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



Notes: The figure shows the average estimated risk-adjusted share of unplanned readmissions over the initial 30 days of the stay. The figure shows the trends pre-2015 (left half of figure) and post-2015 (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. For this reason, the estimated change from 2015–2016 is not economically meaningful. The dollar amounts are adjusted for economy-wide inflation using the aggregate PCE price index.

Figure A5. Risk-Adjusted At Home by Condition, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)

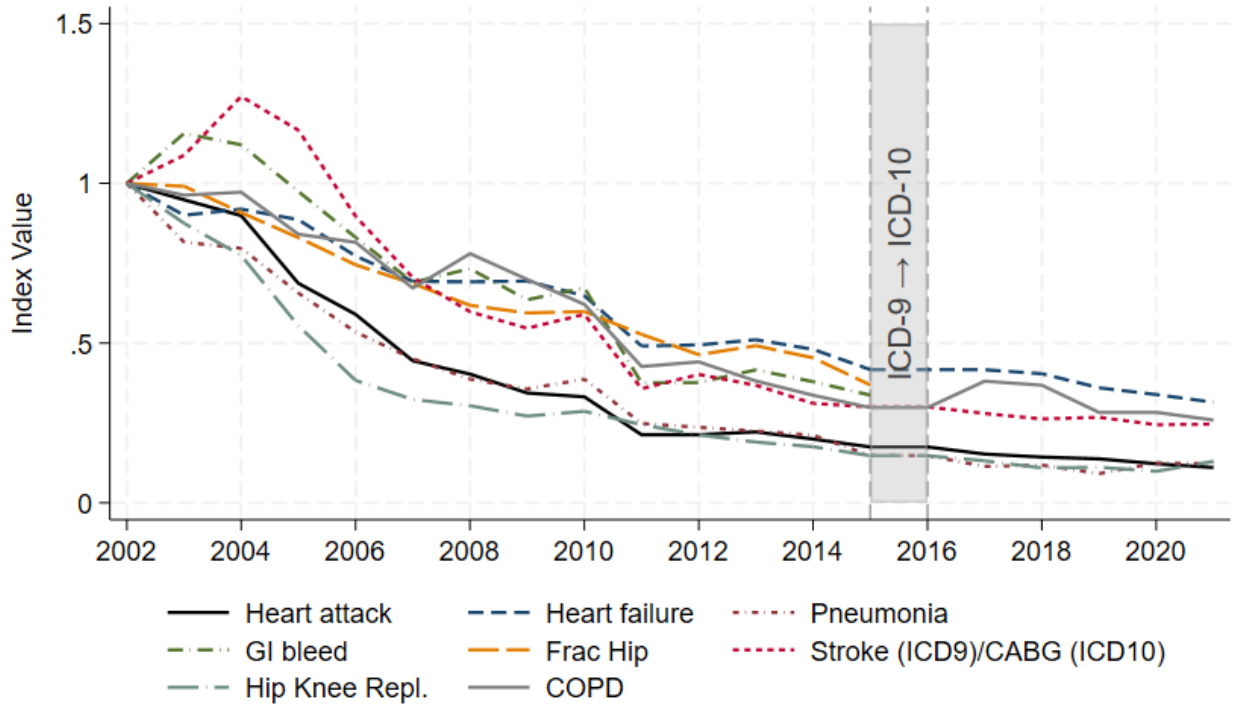


Notes: The figure shows the average estimated risk-adjusted risk-adjusted share of patients returning to home. The figure shows the trends pre-2015 (left half of figure) and post-2015 (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. For this reason, the estimated change from 2015–2016 is not economically meaningful. The dollar amounts are adjusted for economy-wide inflation using the aggregate PCE price index.

Appendix D Robustness Results

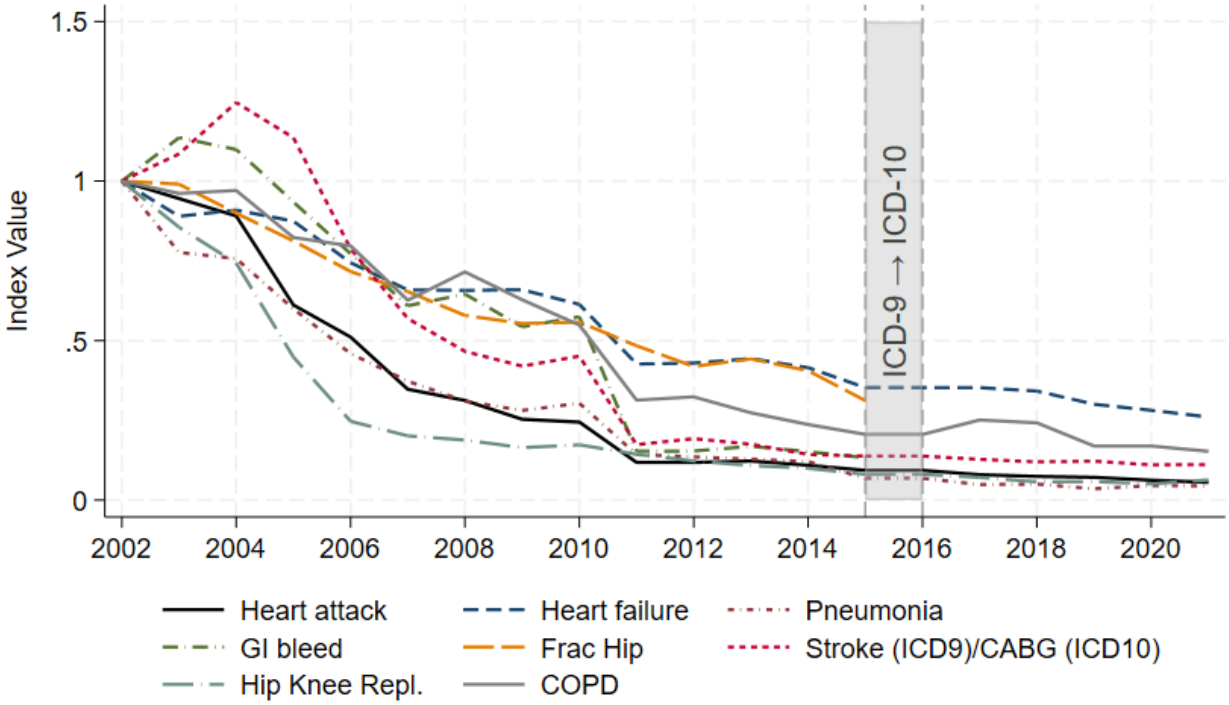
The robustness section of the main text describes the additional analysis presented here.

Figure A6. Quality-Adjusted Price Indexes ($VQALY 100,000), 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



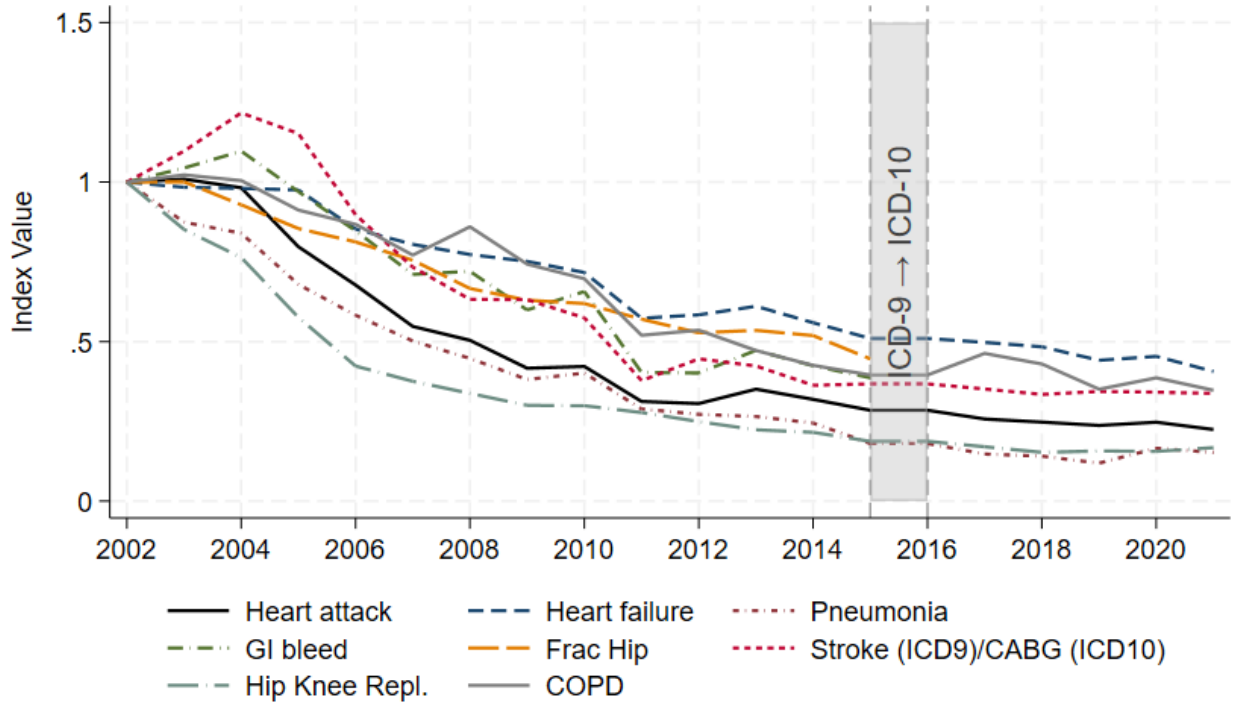
Notes: This figure shows the risk-adjusted chained productivity index. The figure shows the trends pre-2015 (left half of figure) and post-2015 (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. The index is held constant for 2015–2016. The dollar amounts are adjusted for economy-wide inflation.

Figure A7. Quality-Adjusted Cost Index Adjusted Using Laspeyres Formula and Change in Lifetime QALY ($VQALY 100,000), 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



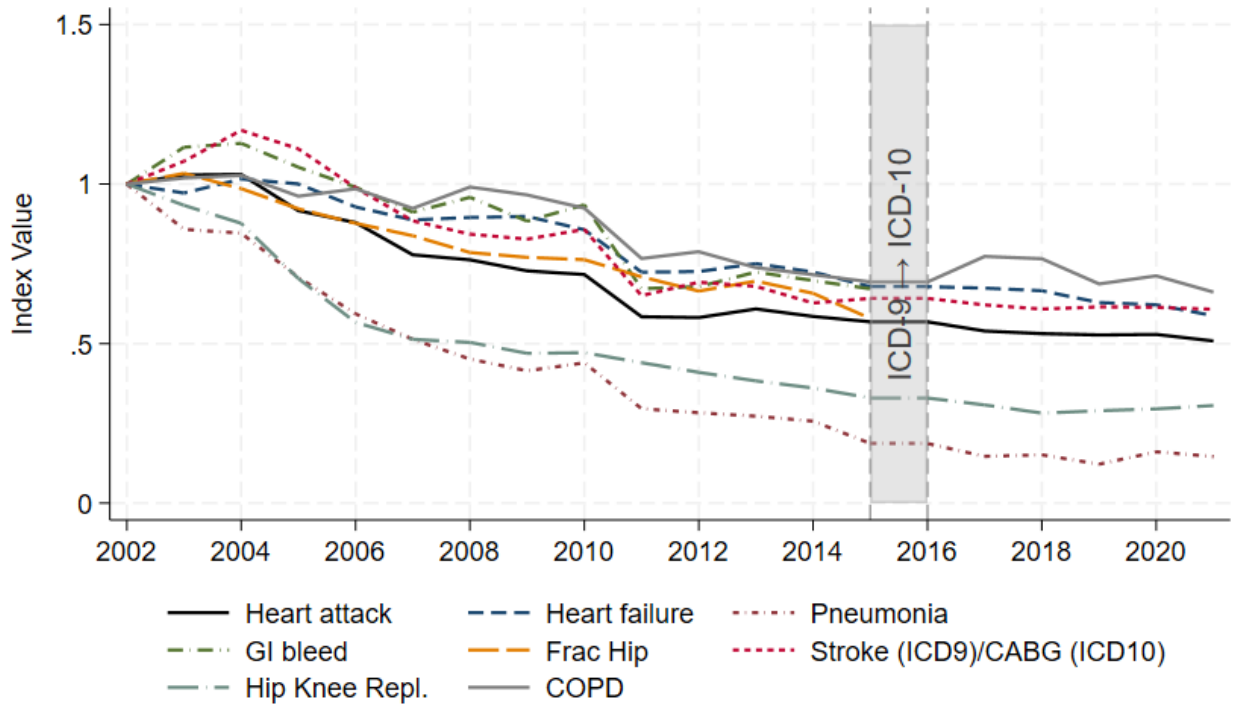
Notes: This figure shows the risk-adjusted chained productivity index. The chaining uses the Laspeyres index formula, rather than the Paasche. The figure shows the trends pre-2015 (left half of figure) and post-2015 (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. The index is held constant for 2015–2016. The dollar amounts are adjusted for economy-wide inflation.

Figure A8. Quality-Adjusted Cost Index Holding Post 90-Day Outcome Fixed ($VQALY 100,000), 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



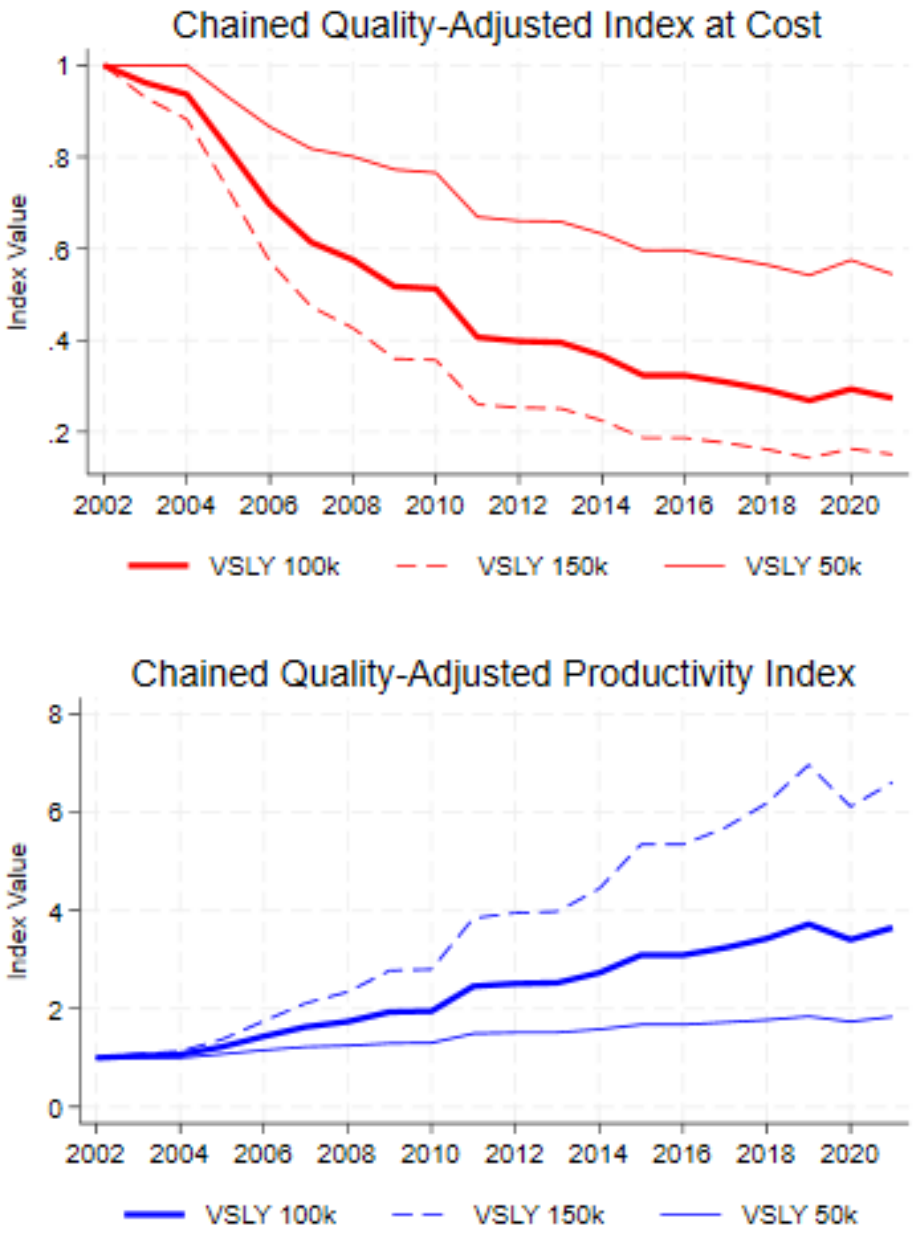
Notes: This figure shows the chained quality-adjusted cost index. The quality metric holds the post-90-day outcome fixed for all observations. The chaining uses the Laspeyres index formula, rather than the Paasche. The figure shows the trends pre-2015 (left half of figure) and post-2015 (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. The index is held constant for 2015-2016. The dollar amounts are adjusted for economy-wide inflation.

Figure A9. Quality-Adjusted Cost Index Using Estimated Lifetime costs ($VQALY 100,000), 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



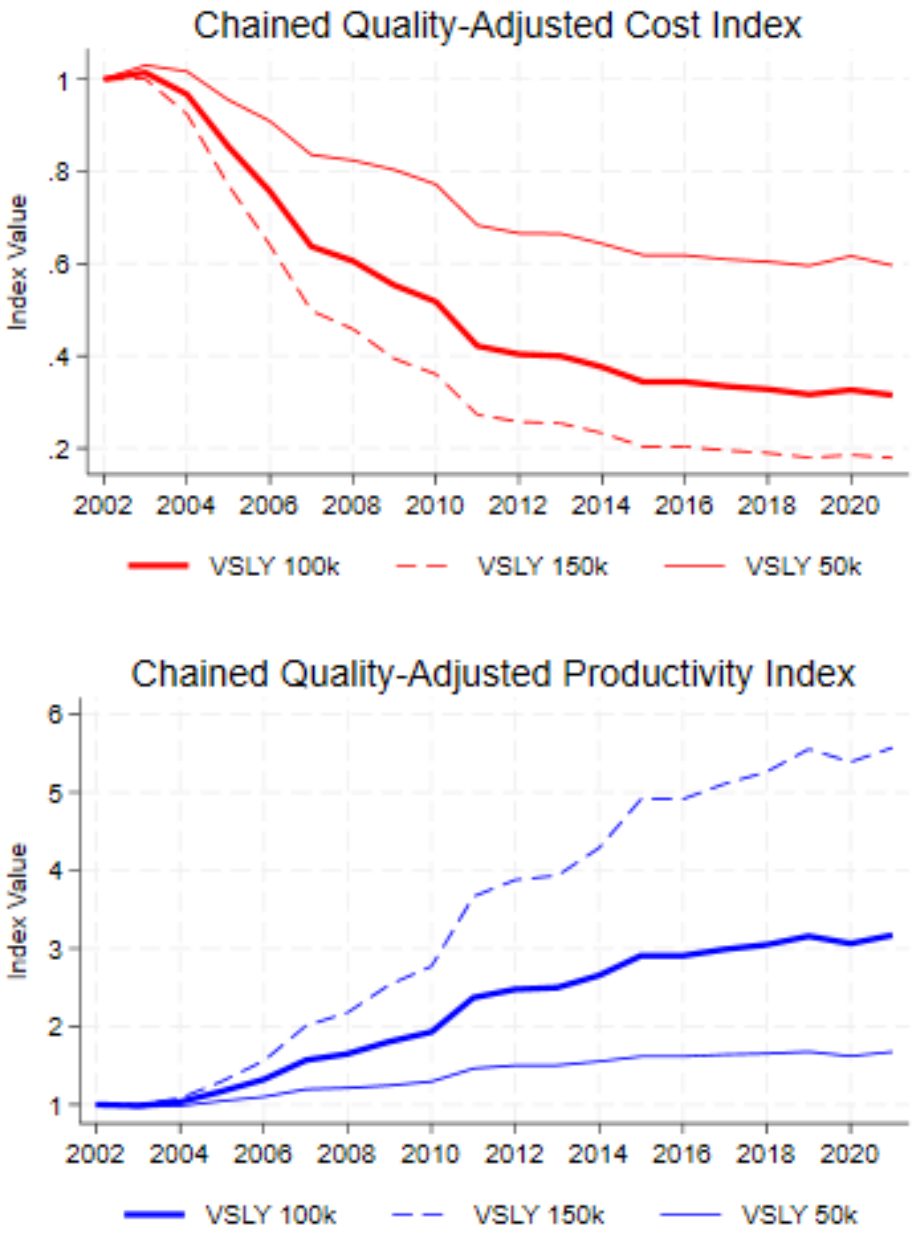
Notes: This figure shows the chained quality-adjusted cost index. Estimated lifetime costs are computed rather than 90-day costs as described in the robustness section of the paper using information in Table A5. The figure shows the trends pre-2015 (left half of figure) and post-2015 (right half of figure). The gray area indicates the transition from disease coding using ICD-9 to coding using ICD-10. The index is held constant for 2015–2016. The dollar amounts are adjusted for economy-wide inflation.

Figure A10. Aggregate Chained Paasche Quality-Adjusted Cost and Productivity Holding Post 90-Day Outcome Fixed, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



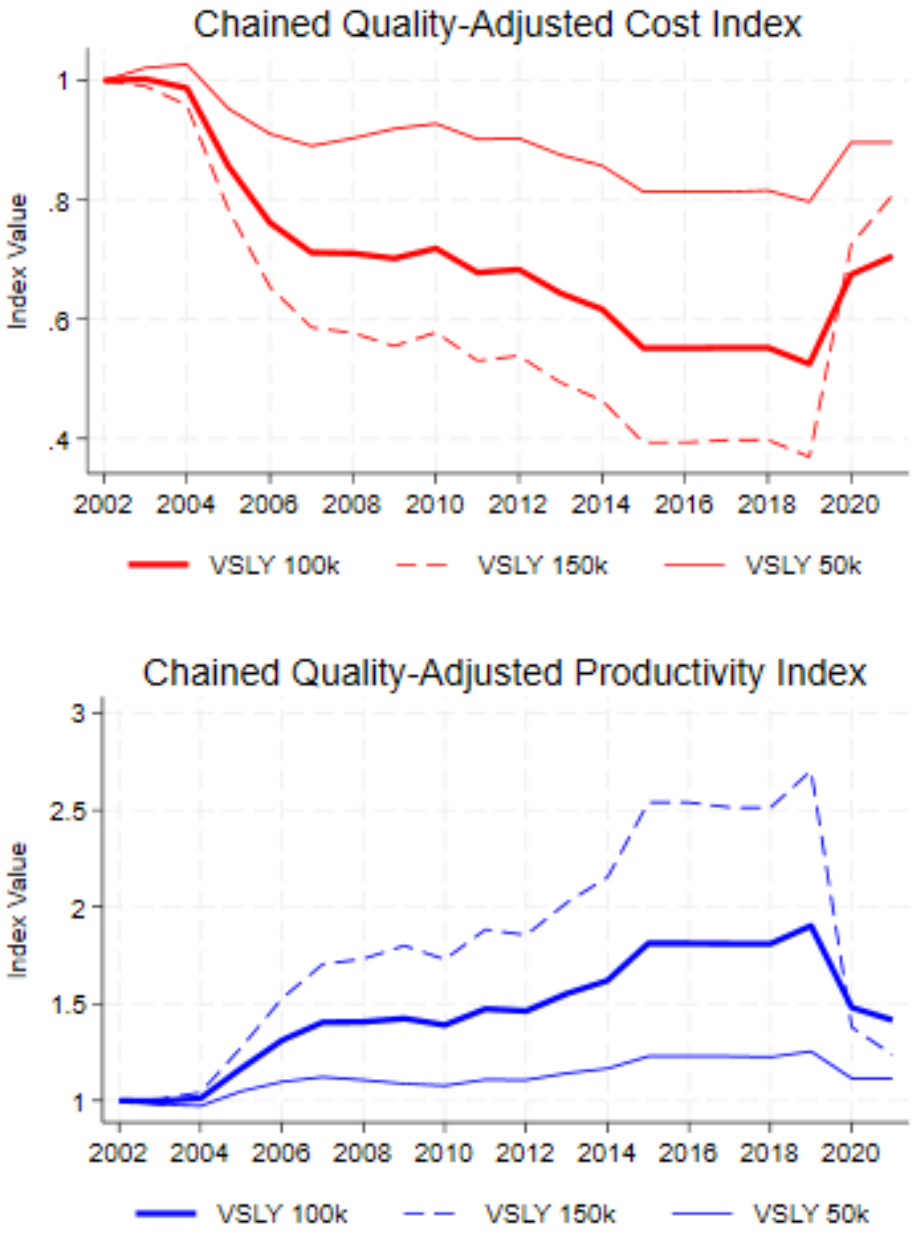
Notes: Chained measures applying $100,000 VSLY, $50,000 VSLY, and $150,000. The top panel shows the quality-adjusted cost index, and the bottom panel shows the productivity measure. Index chaining is weighted by the expenditure share of each condition in the prior period. Note that different conditions are used in the chaining in the pre-2015 period and the post-2015 period due to the ICD-9 and ICD-10 change. The index is held at no change from 2015 to 2016 for those conditions observed in pre- and post- period.

Figure A11. Aggregate Chained Paasche Quality-Adjusted Cost and Productivity Applying Lifetime Cost Estimate, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



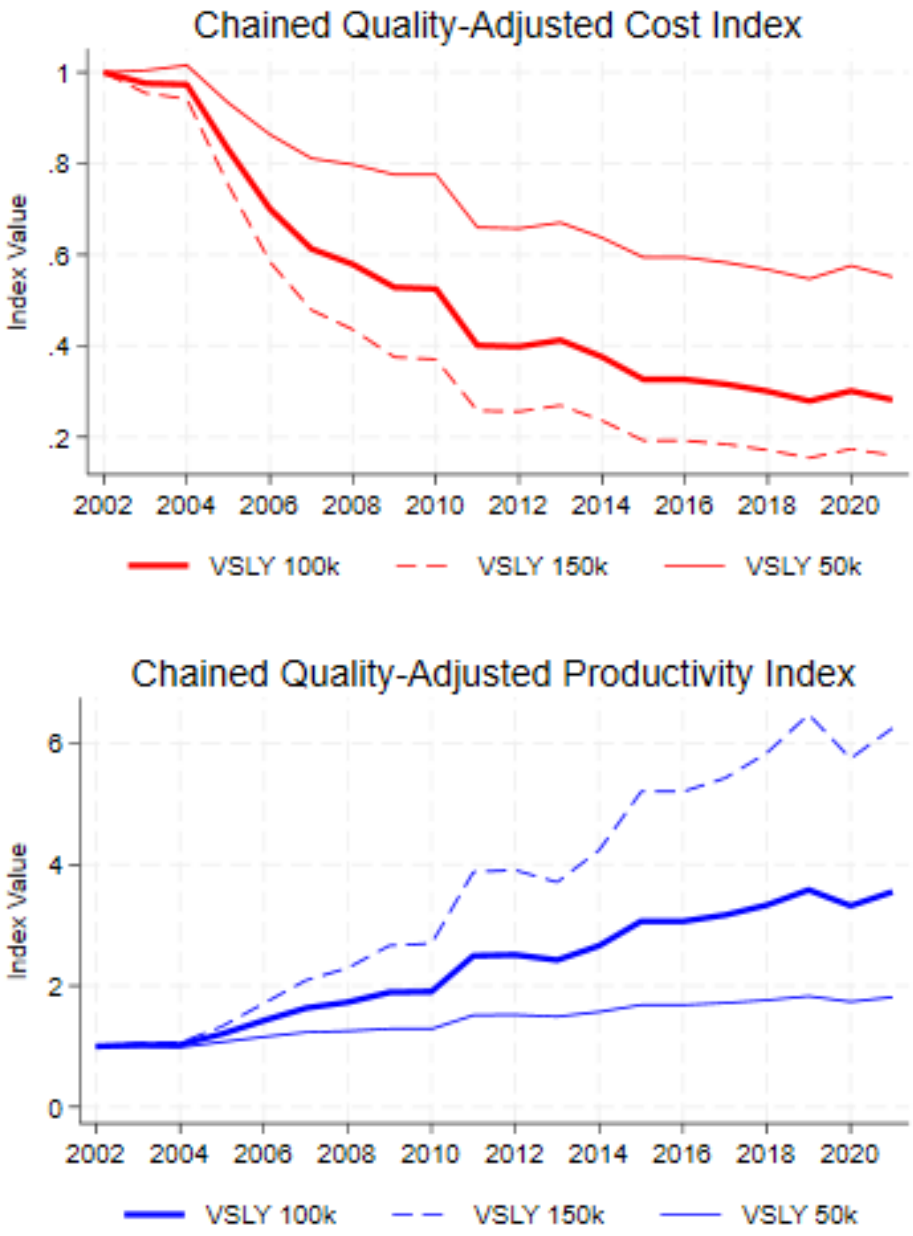
Notes: Chained measures applying $100,000 VSLY, $50,000 VSLY, and $150,000. The top panel shows the quality-adjusted cost index, and the bottom panel shows the productivity measure. Index chaining is weighted by the expenditure share of each condition in the prior period. Note that different conditions are used in the chaining in the pre-2015 period and the post-2015 period due to the ICD-9 and ICD-10 change. The index is held at no change from 2015 to 2016 for those conditions observed in pre- and post- period.

Figure A12. Aggregate Chained Quality-Adjusted Cost and Productivity Using Only Demographic Adjustment and No Adjustment Based on Diagnosis Codes, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)



Notes: Chained measures applying $100,000 VSLY, $50,000 VSLY, and $150,000. The top panel shows the quality-adjusted cost index, and the bottom panel shows the productivity measure. Index chaining is weighted by the expenditure share of each condition in the prior period. Note that different conditions are used in the chaining in the pre-2015 period and the post-2015 period due to the ICD-9 and ICD-10 change. The index is held at no change from 2015 to 2016 for those conditions observed in pre- and post- period.

Figure A13. Aggregate Chained Quality-Adjusted Cost and Productivity Using Additional Hospital Controls to Adjust Outcomes, 2002–2015 (ICD-9) and 2016–2021 (ICD–10)



Notes: Chained measures applying $100,000 VSLY, $50,000 VSLY, and $150,000. The top panel shows the quality-adjusted cost index, and the bottom panel shows the productivity measure. Index chaining is weighted by the expenditure share of each condition in the prior period. Note that different conditions are used in the chaining in the pre-2015 period and the post-2015 period due to the ICD-9 and ICD-10 change. The index is held at no change from 2015 to 2016 for those conditions observed in pre- and post- period.

Table A2. Quality-Adjusted Cost Indexes by Condition, VSLY $100,000, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)

	Condition							
	Heart attack	Heart failure	Pneumonia	GI bleed	Frac Hip	Stroke/CABG	Hip Knee Repl.	COPD
Year								
2002	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2003	0.992	0.938	0.853	1.198	1.020	1.083	0.892	0.998
2004	0.955	0.971	0.840	1.173	0.939	1.270	0.803	1.006
2005	0.741	0.933	0.698	1.031	0.852	1.163	0.586	0.892
2006	0.659	0.822	0.581	0.903	0.775	0.900	0.423	0.877
2007	0.519	0.756	0.500	0.754	0.715	0.711	0.364	0.747
2008	0.489	0.753	0.437	0.804	0.640	0.619	0.348	0.856
2009	0.417	0.740	0.399	0.685	0.605	0.565	0.309	0.763
2010	0.409	0.692	0.424	0.729	0.598	0.605	0.316	0.686
2011	0.284	0.541	0.282	0.425	0.536	0.372	0.281	0.486
2012	0.285	0.548	0.269	0.433	0.485	0.427	0.253	0.510
2013	0.304	0.573	0.259	0.483	0.520	0.401	0.229	0.447
2014	0.281	0.539	0.243	0.440	0.481	0.342	0.213	0.398
2015	0.254	0.480	0.175	0.402	0.396	0.345	0.185	0.361
2016	0.254	0.480	0.175			0.345	0.185	0.361
2017	0.227	0.472	0.136			0.325	0.168	0.452
2018	0.217	0.460	0.141			0.313	0.147	0.437
2019	0.213	0.422	0.112			0.317	0.153	0.344
2020	0.203	0.408	0.150			0.307	0.151	0.351
2021	0.189	0.376	0.136			0.309	0.168	0.310

Notes: The estimates show the risk-adjusted chained cost index for the nine conditions studied. These estimates correspond to the main results presented in Figure 5. Due to the ICD-9 to ICD-10 change from 2015 to 2016, the index is held constant across these years for the remaining conditions.

Table A3. Risk-Adjusted Survival by Condition in QALYs, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)

	Condition							
	Heart attack	Heart failure	Pneumonia	GI bleed	Frac Hip	Stroke/CABG	Hip Knee Repl.	COPD
Year								
2002	4.296	2.446	3.236	4.393	3.126	3.855	8.385	3.447
2003	4.320	2.476	3.290	4.363	3.144	3.846	8.436	3.459
2004	4.355	2.482	3.299	4.377	3.187	3.808	8.483	3.458
2005	4.480	2.500	3.345	4.411	3.236	3.830	8.613	3.489
2006	4.544	2.549	3.408	4.454	3.299	3.929	8.763	3.503
2007	4.658	2.587	3.458	4.513	3.348	4.023	8.824	3.557
2008	4.688	2.597	3.507	4.505	3.418	4.086	8.852	3.527
2009	4.793	2.617	3.549	4.554	3.475	4.134	8.913	3.573
2010	4.796	2.637	3.526	4.549	3.477	4.115	8.897	3.607
2011	4.970	2.723	3.659	4.730	3.533	4.309	8.954	3.707
2012	4.967	2.716	3.673	4.722	3.587	4.267	8.992	3.697
2013	4.949	2.707	3.687	4.703	3.553	4.292	9.028	3.732
2014	4.982	2.733	3.702	4.733	3.593	4.350	9.049	3.768
2015	5.037	2.776	3.815	4.766	3.704	4.363	9.108	3.803
2016	5.328	2.499	3.475			7.050	10.481	3.532
2017	5.392	2.508	3.564			7.103	10.506	3.469
2018	5.422	2.519	3.556			7.118	10.550	3.485
2019	5.438	2.561	3.648			7.128	10.539	3.578
2020	5.489	2.588	3.568			7.186	10.569	3.580
2021	5.529	2.609	3.583			7.144	10.509	3.623

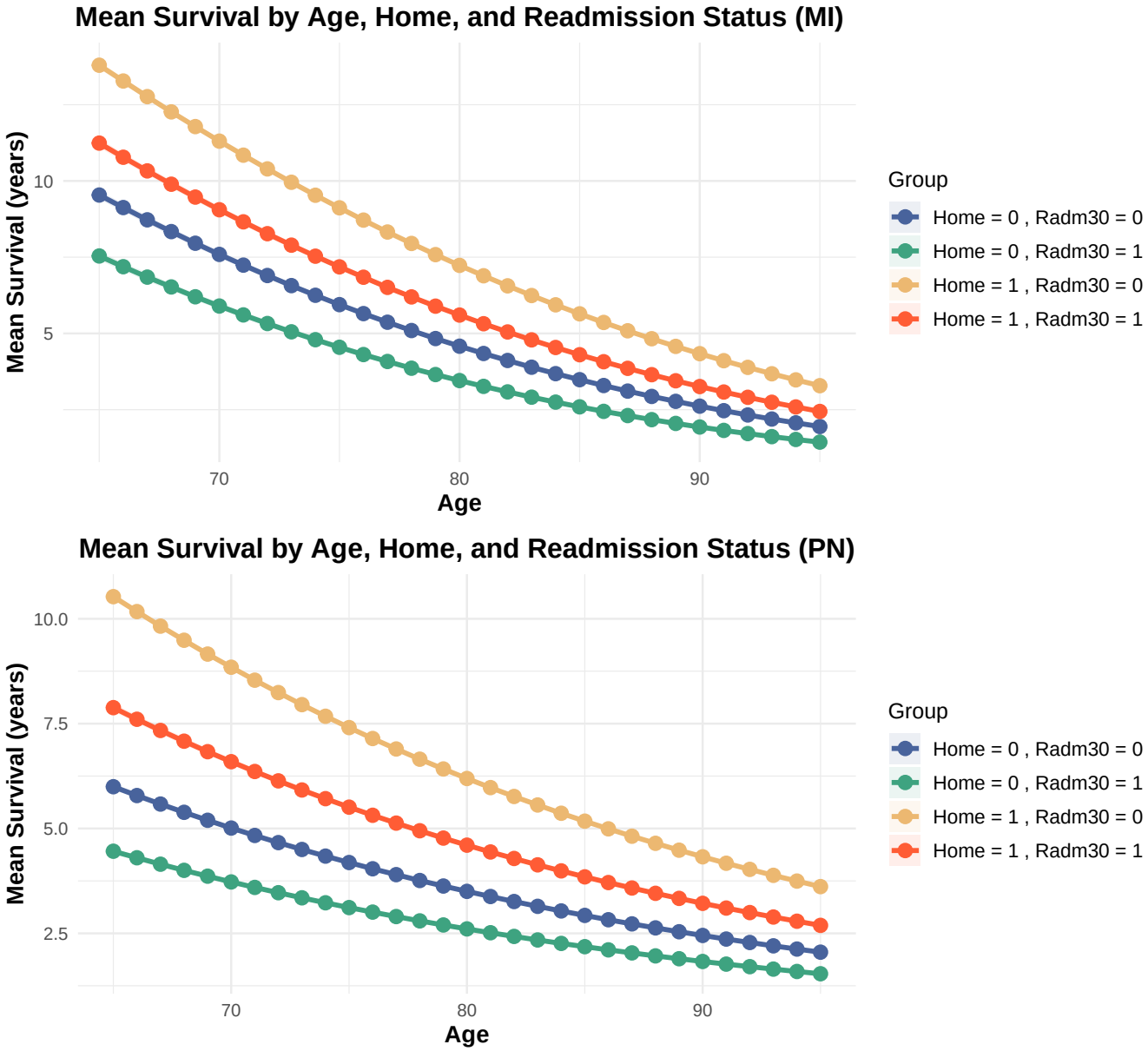
Notes: The estimates show the risk-adjusted lifetime QALY by condition used in our main results in Figure 5.

Table A4. Risk-Adjusted Cost by Condition in Thousands of Dollars, 2002–2015 (ICD-9) and 2016–2021 (ICD-10)

	Condition							
	Heart attack	Heart failure	Pneumonia	GI bleed	Frac Hip	Stroke/CABG	Hip Knee Repl.	COPD
Year								
2002	37.060	28.796	25.201	23.678	36.496	28.341	33.115	25.148
2003	39.145	29.835	26.088	24.750	39.029	29.735	34.127	26.272
2004	41.142	31.520	26.582	25.561	39.890	30.441	34.914	26.394
2005	41.553	32.050	25.920	25.534	40.573	29.844	34.969	26.104
2006	42.653	32.535	26.843	26.138	42.729	30.803	36.075	27.081
2007	42.637	33.411	27.399	26.749	43.890	31.756	36.321	27.624
2008	42.959	34.264	28.195	27.590	45.543	33.105	37.428	28.291
2009	45.639	35.672	29.588	27.746	48.457	34.634	38.672	29.320
2010	45.022	35.229	29.119	28.976	48.078	34.958	37.757	29.444
2011	43.416	34.204	28.158	27.394	48.090	33.493	38.748	27.907
2012	43.176	33.972	28.208	27.114	48.432	33.656	38.347	28.218
2013	44.190	34.631	28.420	28.111	48.265	33.914	37.936	27.795
2014	43.799	34.995	28.116	28.379	48.385	33.812	37.191	28.036
2015	44.644	35.012	28.371	28.900	49.043	35.464	37.514	28.515
2016	51.529	41.867	35.786			84.209	32.567	35.019
2017	51.730	42.160	34.758			84.319	31.900	35.895
2018	52.228	42.155	35.172			82.704	31.730	36.250
2019	52.932	42.427	35.448			84.821	31.862	35.866
2020	55.170	43.656	36.550			87.633	34.330	36.854
2021	55.097	42.237	34.453			84.107	31.662	36.271

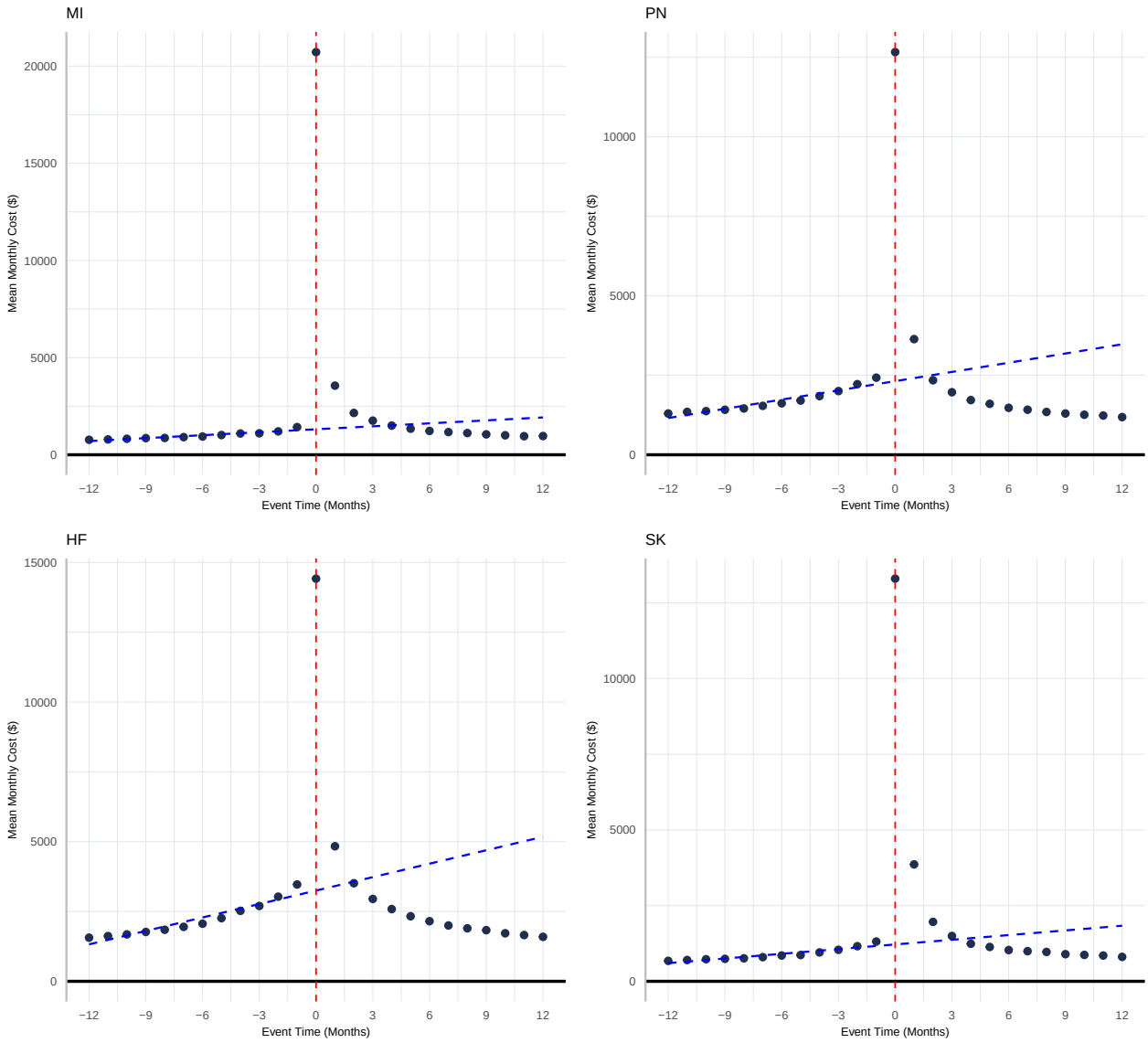
Notes: The estimates show the risk-adjusted 90-day cost by condition used in our main results in Figure 5.

Figure A14. Estimated Survival by Age, Readmission, and Discharged Home for Heart Attacks (also known as Myocardial Infarction (MI)) and Pneumonia (PN)



Notes: This figure displays predicted survival after discharge for various age categories for two conditions, heart attacks (MI) in the upper panel and pneumonia (PN) in the lower panel. The survival model uses the patient age, whether the patient returned to home after the 90-day period and whether the patient experienced a readmission. For example, a 70-year-old patient heart attack patient that does not return to home and experiences a readmission has a survival of about 6 years, while a 70-year-old patient that returns to home without a readmission is estimated to survive about 11 years.

Figure A15. Cost Pre-And Post Episode Heart Attack (AI), Pneumonia (PN), Heart Failure (HF) and Stroke (Sk)



Notes: The figure shows the cost around the months of the indexed admission. The time in months around the admission that starts the episode is on the x-axis with a mark of 0 corresponding to the month when the episode occurred. The black dots show the monthly average cost around the admission, while the blue doted line graphs the pre-trend. The figure clearly shows a sharp drop in expenditures after the first month that falls at or below trend after 3 months, providing evidence in support of a 90-day window.

Table A5. Costs Pre-, During-, and Post-Event, and Cost Share (Annual Cost Post / 90-Day Episode Cost)

Condition	Pre-Period	Event Period	Post-Period	Post–Pre	(Post–Pre)×12	Share
CABG	652.08	21,572.85	1,093.93	441.85	5,302.21	0.08
COPD	1,663.85	6,291.35	2,097.53	433.67	5,204.10	0.28
Hip fracture	826.55	9,082.00	1,024.55	198.00	2,376.05	0.09
GI bleed	1,172.19	5,452.81	1,385.36	213.16	2,557.98	0.16
Heart attack	850.40	8,810.01	1,207.82	357.41	4,288.97	0.16
Heart failure	1,784.14	7,588.73	2,071.36	287.22	3,446.69	0.15
Pneumonia	1,435.24	6,215.63	1,450.38	15.14	181.70	0.01
Stroke	751.38	6,374.20	1,029.53	278.15	3,337.80	0.17
Hip/knee replacement	778.19	8,302.49	873.24	95.05	1,140.61	0.05

Notes: Share = (Post–Pre)×12 / (Event Period × 3). This table contains the figures used in our lifetime cost robustness check. Figure A15 suggests that post 90-day costs are minimal as costs fall below pre-trend after the 90-day period. This table constructs an alternative lifetime cost estimate assuming there is no trend in cost using monthly average cost estimates of patients. To do this, we estimate monthly average cost for the pre-period months 6 to 12 months prior to the episode (column 1). We then construct average monthly cost during the episode (column 2), and 9 months after the episode (column 3). We estimate the additional monthly cost to be post-period minus pre-period (column 4), and the annualized cost (column 5). Finally, we construct the cost share (column 6) as the share of annual additional cost divided by the 90-day costs. We then multiply the share by the 90-day cost and also by life expectancy to obtain the lifetime cost estimates in our robustness exercise.