

NBER WORKING PAPER SERIES

A SATELLITE ACCOUNT FOR HEALTH IN THE UNITED STATES

David M. Cutler
Kaushik Ghosh
Kassandra Messer
Trivellore Raghunathan
Allison B. Rosen
Susan T. Stewart

Working Paper 27848
<http://www.nber.org/papers/w27848>

NATIONAL BUREAU OF ECONOMIC RESEARCH

1050 Massachusetts Avenue
Cambridge, MA 02138

September 2020. Revised September 2021

We are grateful to Patricia Berglund, Irina Bondarenko, Marcelo Coca Perrillon, Paul Imbriano, Rebecca Woodward, and James Schafer for research assistance, and to participants in our National Health Accounts Advisory Group and the Harvard University Interfaculty Program for Health Systems Improvement for initial advice and support. We thank Corby Garner, Joe Newhouse, Louise Sheiner, the editors, and three anonymous reviewers for very helpful comments, and the National Institute On Aging of the National Institutes of Health for research support under award numbers R37AG047312 and P01AG31098. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research or the National Institute On Aging. The code for replicability purposes has been deposited in the AEA Data and Code Repository, openicpsr-143521. This research was approved by the Institutional Review Board at Harvard University (ID: HUM00096289) and NBER (ID: 16-264). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research or the National Institute On Aging.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by David M. Cutler, Kaushik Ghosh, Kassandra Messer, Trivellore Raghunathan, Allison B. Rosen, and Susan T. Stewart. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

A Satellite Account for Health in the United States

David M. Cutler, Kaushik Ghosh, Kassandra Messer, Trivellore Raghunathan, Allison B. Rosen, and Susan T. Stewart

NBER Working Paper No. 27848

September 2020. Revised September 2021

JEL No. E2,I10,I11,I12,I38

ABSTRACT

This paper develops a satellite account for the US health sector and measures productivity growth in health care for the elderly population between 1999 and 2012. We measure the change in medical spending and health outcomes for a comprehensive set of 80 conditions. Medical care has positive productivity growth over the time period, with aggregate productivity growth of 1.5% per year. However, there is significant heterogeneity in productivity growth. Care for cardiovascular disease has had very high productivity growth. In contrast, care for people with musculoskeletal conditions has been costly but has not led to improved outcomes.

David M. Cutler
Department of Economics
Harvard University
1875 Cambridge Street
Cambridge, MA 02138
and NBER
dcutler@harvard.edu

Kaushik Ghosh
NBER
1050 Massachusetts Ave.
Cambridge, MA 02138
ghoshk@nber.org

Kassandra Messer
University of Michigan
SRC - Health and Retirement Study
3467 ISR-Thompson
Ann Arbor, MI 48106
kasey@umich.edu

Trivellore Raghunathan
University of Michigan
SPH II
1415 Washington Heights
Ann Arbor, MI 48109
teraghu@umich.edu

Allison B. Rosen
Department of Quantitative Health Sciences
University of Massachusetts Medical School
368 Plantation Street, AS9-1083
Worcester, MA 01605
Arosen006@gmail.com

Susan T. Stewart
NBER
1050 Massachusetts Ave
Cambridge, MA 02138
sstewart@nber.org

A online appendix is available at <http://www.nber.org/data-appendix/w27848>

“The welfare of a nation can scarcely be inferred from a measure of national income.”

- Simon Kuznets (1934, p. 7)

Estimating medical care productivity growth is a central economic challenge. The National Income and Product Accounts (NIPAs) report that multi-factor productivity in the health care and social assistance sector declined by 0.4% annually between 1987 and 2018,¹ in contrast to the 0.9% increase in the economy as a whole. Given the size of health care in the economy, about 18% of GDP, low productivity growth in medical care influences aggregate productivity growth. Further, the low rate of measured productivity growth contributes to ongoing concerns about wasted resources in medicine (Chandra and Skinner, 2012).

However, official medical care productivity growth estimates suffer from a variety of problems (Advisory Commission to Study the Consumer Price Index, 1996; Berndt et al., 2001; Dunn, Rittmueller and Whitmire, 2015). Quality adjustment is an often-noted issue. The primary output of medical care is improved health, yet health is not easily measured or attributed to medical interventions. Thus, health improvements are undercounted in official productivity growth estimates. This limitation was evident as far back as the 1930s, as the quotation from Simon Kuznets attests. In addition, the national accounts have difficulties when care from one type of provider substitutes for care from another. Each provider type is treated as a separate industry in the national accounts, for example pharmaceuticals, hospitals, and physicians. Thus, if a medication is developed which avoids the need for hospitalizations, the savings in hospital costs are not credited to the pharmaceutical industry. For each of these reasons, medical care productivity growth by sector may be significantly misstated.

To address these issues, proposals have periodically been made to establish a ‘satellite account’ for health. Such an account would measure health outcomes and medical inputs, and thus more accurately account for new goods and substitution of treatments (Schultze and Mackie, 2002; Abraham and Mackie, 2005; Rosen and Cutler, 2007; Rosen and Cutler, 2009; National Research

¹ This includes health care and social assistance, although social assistance represents less than 10% of the total. Health care services includes inpatient and outpatient facilities but excludes pharmaceuticals and medical devices. Over the 1987-2016 time period, the reported MFP of the pharmaceutical industry was -1.7% annually and the reported productivity growth of the medical equipment and supplies industry was 0.7% annually.

Council, 2010).² The goal of this paper is to develop a satellite account for the US health sector and measure the change in the productivity of medical care over time.

The key design innovation in our accounts is to use medical conditions as the industries, not providers of care. For example, there is an industry for heart disease and a second one for lung cancer. Our accounts make no distinction based on the type of care provided; all that matters to people is how much they spend on care and their resulting health.³

The data needs for a satellite account are substantial, including measures of population health and medical spending matched to clinical characteristics and other sociodemographic factors. We benchmark medical spending and mortality to national data. Our empirical work adds measures of quality of life and allocates health and spending changes to conditions. The primary set we employ is the Medicare Current Beneficiary Survey (MCBS), an ongoing survey of the elderly population with data from 1999–2012. Thus, this initial satellite account is focused on the elderly, who account for one-third of total medical spending.

We start by identifying a set of 80 conditions that are prevalent and severe enough to affect the elderly population. Of these conditions, 78 are clinical, for example heart attack and lung cancer. The remaining two are self-reported moderate and severe frailty (defined below). Frailty is not a disease that clinicians code, but we show that it is vital for both spending and health. We measure the prevalence of each condition using information from medical claims combined with imputations based on self-reports in the National Health and Nutrition Examination Survey (NHANES).

Aggregate medical spending is straightforward to measure. Real spending (\$2010) on the elderly increased by roughly \$5,000 per person over this time period. The change in health is more difficult to ascertain. The concept we wish to measure is quality-adjusted life expectancy (QALE), defined as the expected number of years a person will live, adjusted for the health-related quality of those years. To match our focus on the elderly, we measure QALE as of age 65. Life tables giving mortality by age are readily available from Vital Statistics. We develop an empirical measure of health-related quality of life based on survey data asking about difficulty with

² Health is not the only area with these concerns. Other areas for which satellite accounts are proposed include home production (Landefeld and McCulla, 2000), the environment (Muller et al., 2011), education (Fraumeni et al., 2017), and free goods (Brynjolfsson et al., 2019).

³ For policy purposes, it is likely helpful to have a combination of condition and provider-based accounts. For this reason, and because our analysis is limited to the elderly population, we view the accounts as a complement to existing NIPAs, rather than a substitute.

necessary physical activities, sensory impairments, and how much health interferes with social interactions. These domains capture many aspects of health, but not all; pain, cognition, and mental health symptoms are not asked about directly.⁴ We estimate that over the 1999–2012 time period, quality-adjusted life expectancy increased by about one year at age 65.

The bulk of our empirical analysis estimates how much of the change in spending and health outcomes result from medical treatment changes. The primary assumption we make in our analyses is that in most cases, medical care for a condition affects the health of people with that condition but not the prevalence of other conditions. That is, care for a heart attack influences the length and quality of life for people with a heart attack, but not whether the person develops cancer. The exceptions are medical risk factors such as hypertension and high cholesterol, where significant evidence shows an impact on downstream conditions. The assumption implies that the productivity of medical care can largely be determined by the health and spending trends for people once they have acquired conditions.

Empirical estimation of condition-specific health and spending trends is complex because of comorbidities. If a person has a heart attack and dies, how much of the death should be attributed to the heart attack, and how much to the mental illness which made taking preventive measures difficult? Our method to form condition-specific health and medical spending uses propensity score models to compare people with each condition to a similar group without the condition. The health and spending difference between these two groups is an estimate of the impact of that condition on the relevant outcome. We scale these estimates so that the total adds to national data on mortality and spending. We then transfer a portion of spending and health decrements from direct conditions to medical risk factors. The result is spending and health decrements for each condition in each year, which we use to form productivity growth at the condition level and, by adding up, for medical care as a whole.

Our analysis yields several conclusions. Over this time period we examine, productivity growth in medical care as a whole is relatively rapid: 1.5% annually. This is in contrast to official data, which show virtually no change in productivity over this interval. The high rate of productivity growth is a reflection of the large increase in QALE for the elderly population. As noted above, QALE at age 65 rose by 1 year. However, medical treatment changes accounted for

⁴ We show using other data that the measures we have are highly correlated with these omitted domains of health. Still, this adds an element of caution to our findings.

an even greater increase, roughly 1.7 years, as overall QALE was held back by obesity and other non-medical factors. The value of health gains is large enough to outweigh the increase in lifetime medical spending.

Care for cardiovascular disease showed the largest productivity growth. Cardiovascular disease risk factors are better controlled than they were, and people who have had cardiovascular events live longer with fewer recurrent acute events than they did in the past. This survival gain has come at minimal cost increase, leading to large productivity growth. We validate our findings on cardiovascular disease by comparing our mortality trends to Vital Statistics tabulations and to epidemiological models of cardiovascular disease mortality. Our results match well those using other methodologies.

In contrast to the clear findings for cardiovascular disease, there is considerable heterogeneity in productivity change for other conditions. Musculoskeletal disease is a prototypical example; spending increased rapidly yet survival and quality of life for people with musculoskeletal disorders were barely affected. The same is true for mental illness, although this result is somewhat more tentative, as quality of life measurement for mental illness is less exact. Thus, musculoskeletal disorders and mental health are estimated to have low productivity growth. Overall, the heterogeneity in productivity growth across conditions suggests a significant opportunity for productivity-enhancing reforms in health care.

The paper is structured as follows. The first section discusses the conceptual issues in health accounting. The second section describes our data. The third section details our methodology for estimating condition-specific spending growth and health improvement. The fourth and fifth sections estimate spending changes and health improvement by condition. The sixth section shows productivity growth. The last section concludes.

I. Accounting for Health

A. Current Treatment of Medical Care in NIPAs and Problems

The United States has two sets of medical spending accounts, one a part of the NIPAs and a second maintained by the Centers for Medicare and Medicaid Services (CMS). The two have many features in common (Sensenig and Wilcox, 2001). Both measure medical spending, not health outcomes. Both divide spending into industries based on the type of provider, not the disease. In the NIPAs, there are 13 industries relevant to health care, reflecting a mix of goods

(prescription drugs, nonprescription drugs, and medical equipment) and services (e.g., hospital care and physician care).⁵

The Bureau of Labor Statistics (BLS) measures prices in each industry via surveys. Providers in the price surveys are asked to price the same bill over time, holding constant patient characteristics, payment source, and method of payment (for example, a hospitalization for a heart attack without significant comorbidities, covered by the Blue Cross-Blue Shield HMO plan). As in most industries, quality adjustment is designed to be incorporated at the price index stage. Real output is nominal spending deflated by quality-adjusted prices (BLS, 2007).

In practice, there are several difficulties with the national account treatment of health. We demonstrate these issues with an example consisting of a single condition, heart disease. The first problem is adjusting for changing quality. Imagine that a new surgical procedure is developed that allows people with heart disease to live longer. The impact of the new surgery on the medical price index will depend on how it is paid for. If the new service is placed in its own payment bucket, it will be linked into the existing price index and its introduction alone will not affect price trends. If the new surgery is placed in the same billing category as an older surgery or is put there by BLS,⁶ the price index will rise or fall depending on the relative reimbursement for the old and new surgery. However, neither of these methods has an explicit quality adjustment.

BLS does not have a good quality adjustment in health care. Respondents to the BLS pricing surveys are asked to indicate if the good has changed materially, but few say so. And many changes in quality result from a series of small steps more than one large change. This is true in many industries, so other quality-adjustment techniques are standard. In many industries, quality adjustment is done through hedonic analysis (Groshen et al., 2017). For example, prices for each procedure or medication over time could be regressed on characteristics of the therapy, including outcomes and side effects (Cockburn and Anis, 2001; Comanor, Schweitzer and Carter, 2005), with the non-quality component of a price change being the price index.

For several reasons, however, hedonic analysis is unlikely to work well in medical care (Berndt et al., 2000): consumers rarely pay the full cost of care they use, so marginal valuations

⁵ The full list of services includes professional services (physicians' services, dental services, eyeglasses and eye care, and other medical professionals) and hospital and related services (hospital services, and nursing homes and adult day care). Health insurance is its own category. CMS further decomposes spending by payer (public insurance, private insurance, out-of-pocket, etc.), though that is not our concern in this paper.

⁶ For example, BLS made brand and generic drugs equivalent in the medical spending account.

are not equal to price; hedonics do not work well in life-or-death decisions; information asymmetries mean that consumers are unlikely to know the real value of services when they receive them; and health outcomes are difficult to parcel out to particular interventions, making it difficult for doctors and patients to know the consequences of what they are doing. As a result, hedonic analysis is not used in health care.

The impact of omitting quality change could be substantial. Life expectancy at birth increased 3.2 years between 1990 and 2017, at a time when medical care productivity was falling. If even a portion of the health improvement were due to medical care and this effect was captured, the overall evaluation of medical care productivity growth could change considerably.

A second problem with current methods is that they do not account for substitution across industries. Returning to the heart disease example, suppose that a new class of pharmaceuticals is developed that substitutes for surgery in people with heart disease. No lives are saved, but the new medication is sufficiently inexpensive that less is spent in total. Regardless of whether the new drug is linked into the existing pharmaceutical index or included in a basket with other medications, the national accounts will not show any productivity gain from reduced surgery. That is because pharmaceuticals and hospital care are different industries. In essence, the shift from surgery to medication will be attributed to consumer tastes, not to productivity growth.

To make progress on both of these issues, we need to partition medical care into more appropriate industries. The assumption we make is that the patient cares about their health and the total amount they spend on treatment, but not what treatment they receive. That is, pills and surgery differ only in their cost and whether their health outcomes. As a result, the relevant industries are not the providers of therapy but rather the conditions being treated.⁷

The Bureau of Economic Analysis has reached a similar conclusion and has recently started experimentally measuring medical spending on a condition basis (Dunn et al., 2015). However, the methods to do this are complex, as we explain below. And no attempt has been made to measure health outcomes systematically.

B. A Satellite Health Account

⁷ For some purposes, one may care about conventional industries as well, for example in determining whether costs for particular inputs are increasing too much. Thus, the conventional accounts have complementary value to what we propose.

We start with a conceptual description of a satellite health account. The schematic for such an account is depicted in Figure 1. Health inputs include medical care, non-medical consumption and behaviors (e.g., cigarettes, physical activity, overall caloric intake), the environment, and genetics. We have data on the first two but leave aside environmental and genetic factors. The primary outcome of the health sector is the consumption value of better health. There is an investment value as well—higher earnings in the future—but this is not particularly important in the elderly population we analyze.

The middle part of Figure 1 shows the combination of these inputs and outcomes into population trends. Health has both a length-of-life and a quality dimension. We combine them into quality-adjusted life expectancy (QALE), the numbers of years a person can expect to live, adjusted for the quality of those years. Since our data is for the elderly population, we estimate QALE as of age 65. We summarize medical spending the same way, as expected spending from age 65 until death.

Aggregate trends are informative about medical care productivity growth but not definitive. Both health and medical spending change over time for reasons beyond just changes in medical technology – most importantly prevalence changes.⁸ The third part of the analysis decomposes spending and health trends into the productivity of medical care. We do this by estimating the spending and outcomes associated with treating each condition. Before discussing the empirical implementation, we demonstrate the ideas theoretically. The framework follows other work on the productivity of medical care, for example Murphy and Topel (2006).

Individual behavior

Consider a population of *ex-ante* identical individuals who receive utility from two goods: health and non-medical consumption.⁹ Health is denoted $H_t(a)$ for a person aged a in year t . We denominate health so that it ranges from 0 (death) to 1 (perfect health). In any period, utility is multiplicative in health and a concave function of non-medical consumption, x_t : $U_t(a) = H_t(a) U(x_t)$, where $U' > 0$ and $U'' < 0$. Thus, greater non-fatal health increases the value of all other consumption. Lifetime utility is the present value of flow utility for each year the person is alive.

⁸ Nordhaus (2013) and Jones and Klenow (2016) both calculate the value of additional longevity and add it to consumption gains over time. Neither considers quality of life.

⁹ For convenience, we omit labor supply. Adding a labor-leisure choice would not change the results materially.

Denote $S_t(a_0)$ as an indicator for whether the person is alive in year t conditional on being alive at age a_0 in $t=0$. ρ as the discount rate. Lifetime welfare is then given by:

$$(1) \quad W(a_0) = \sum_{t=0}^{\infty} (1+\rho)^{-t} S_t(a_0) H_t(a) U(x_t) .$$

If people are fully insured or nearly so, medical spending in any period will be the average spending in the population, denoted $m_{ave,t}$.¹⁰ The budget constraint is that the present value of medical and non-medical consumption must be less than or equal to the present value of income plus initial assets (A_0). With perfect annuity markets, this will be an equality:

$$(2) \quad \sum_{t=0}^{\infty} (1+\rho)^{-t} S_t(a_0) (Y_t - x_t - m_{ave,t}) + A_0 = 0 .$$

For convenience, we assume the interest rate is equal to the discount rate, but this is not necessary if individuals can borrow and lend. Individuals maximize utility (1) subject to the budget constraint (2). The formulation of this problem is given in equation (3):

$$(3) \quad W(a_0) = \sum_{t=0}^{\infty} (1+\rho)^{-t} \{S_t(a_0) H_t U(x_t) + \mu S_t(a_0) (Y_t - x_t - m_{ave,t})\} + \mu A_0 .$$

Equation (3) is denominated in utils. To translate this into dollars, we divide both sides by the marginal utility of wealth, μ .

There are two objects of choice in equation (3), medical spending and non-medical consumption. While medical care decisions are the outcome of some optimization problem, it is not clear that it is a rational decision. Our empirical work does not rely on the optimality of medical care decision-making. We assume only optimization of non-medical consumption. This implies that the marginal utility of non-medical consumption is the same in all periods; $H_t U'_t = \mu$.

Health and Medical Treatments

We now provide more detail on the medical sector. People have one or more medical conditions, denoted by c . The prevalence of condition c at time t is $prev_{ct}$, and spending is m_{ct} . The vectors of condition prevalence and medical spending at time t are given by \mathbf{prev}_t and \mathbf{m}_t . Average

¹⁰ In the absence of this assumption, the value of medical spending varies because marginal utility of income changes with spending on medical care.

medical spending is the inner product of these two vectors: $\mathbf{m}_{ave,t} = \mathbf{prev}_t \cdot \mathbf{m}_t$.

Mortality for condition c denoted is γ_{cat} , and the vector of mortality rates is $\boldsymbol{\gamma}_{at}$. This varies by age¹¹ and may change over time as medical treatments and other health inputs change. In our framework, as in our data analysis, we assume mortality rates for different conditions are independent—that is, γ_{cat} does not depend of other conditions the person has. This assumption is largely for data reasons; our data do not permit a richer mortality structure. Aggregate mortality is the prevalence-weighted average of the mortality rates for individual conditions: $\Gamma_{at} = \mathbf{prev}_{at} \cdot \boldsymbol{\gamma}_{at}$, and this, in turn, determines survival at each age up to a : $S_t(a) = \prod_{\tau=0}^{t-1} (1 - \Gamma_{a\tau})$, for $\tilde{a} = a_0 + \tau$.

We model health-related quality of life as linear in the decrement associated with each condition. Let \mathbf{h}_{at} denote the vector of health decrements associated with each condition, \mathbf{h}_{cat} . Average quality of life is $H_t(a) = 1 - \mathbf{prev}_{at} \cdot \mathbf{h}_{at}$.

Now consider treatment changes for condition c . Differentiating equation (3) and considering a discrete change in medical care, Δm_c , the change in welfare associated with treatment change is given by:

$$(4) \quad \frac{\Delta W(a_0)}{\mu} = \sum_{t=0}^{\infty} (1+\rho)^{-t} \left\{ \left(\frac{dS_t}{d\gamma_c} \right) \left(\frac{d\gamma_c}{dm_c} \right) \left(\frac{U_t}{U_t} + (Y_t - X_t - m_{ave,t}) \right) - S_t(a) \frac{U_t}{H_t U_t} \left(\frac{dh_c}{dm_c} \right) \right\} \text{prev}_{ct} \Delta m_c \\ + \sum_{t=0}^{\infty} (1+\rho)^{-t} \left\{ \left(\frac{dS_t}{d\text{prev}_k} \right) \left(\frac{U_t}{U_t} + (Y_t - X_t - m_{ave,t}) \right) - S_t(a) \frac{U_t}{H_t U_t} h_{kat} \right\} \left(\frac{d\text{prev}_k}{dm_c} \right) \Delta m_c \\ - \sum_{t=0}^{\infty} (1+\rho)^{-t} S_t(a) \text{prev}_{ct} \Delta m_c$$

The first row is the direct effect on health. It has two components. $\left(\frac{dS_t(a)}{d\gamma_c} \right) \left(\frac{d\gamma_c}{dm_c} \right) \text{prev}_{ct} \Delta m_c$ is the impact of medical care on survival to age a . The welfare value of this is found by multiplying the survival gain by $\left(\frac{U_t}{U_t} + (Y_t - X_t - m_{ave,t}) \right)$, flow utility in year t plus the increment to future consumption. $\left(\frac{dh_c}{dm_c} \right) \text{prev}_{ct} \Delta m_c$ is the impact of medical treatment changes on quality of life. The welfare value of this is found by multiplying by $- S_t \frac{U_t}{H_t U_t}$, the dollar value of improved health among those alive to experience it.

The second row involves clinical risk factors. It is operative when changes in medical care

¹¹ Mortality rates are allowed to vary by age, but spending on each condition is not. This is for data reasons; it is difficult to disaggregate spending by condition, let alone add an age dimension.

for one condition, c , affect the prevalence of another condition, k : $\left(\frac{d\text{prev}_k}{dm_c}\right)$. The welfare impact of such a change is given by the health benefit associated with changes in the prevalence of k , the term in $\{.\}$. We term conditions that affect the prevalence of other conditions risk factors, and those with a direct impact on length or quality of life direct conditions.

The third row is the spending effect. The projected lifetime spending change due to a change in medical treatments is the present value of spending increases by people who are alive at each age.

Equation (4) is the multi-factor productivity (MFP) of medical treatment changes, or what Abraham and Mackie (2006) term ‘social profit.’ MFP can be positive or negative. We express MFP in dollars and as a growth rate, dividing by the net value of medical care in the base period.

Equations (3) and (4) are written from the perspective of the average person in society (a ‘per capita’ measure). An alternative is to estimate productivity growth for a person with a condition (a ‘per case’ measure). The per capita and per case measures are approximately related by the prevalence rate.¹² In evaluating each condition individually, we present the per case metric. That is, we ask the question: how is the welfare of people with condition c affected by treatment changes for that condition, assuming the person had the condition in the first and all subsequent years?¹³ We do this so that we can more easily compare across conditions. In contrast, when we evaluate medical care as a whole, we consider treatment changes per capita: what has been the impact on social welfare of all the medical care changes that have occurred, weighting each by the share of people affected by them?

The primary difficulty with estimating equation (4) empirically is isolating the impact of treatment changes from other factors influencing health and medical spending. We detail our approach to this issue after first describing the data we employ.

II. Data and Population Trends

The data needs for a satellite health account are shown in Figure 1. At the individual level, we need data on health status and medical spending for a representative share of the population. Information on other factors influencing health are needed to separate the role of medical advance from other factors that can drive outcomes.

¹² These two methods are not exactly the same because survival is non-linear in prevalence.

¹³ For risk factors, this includes the effect of having the risk factor on downstream conditions.

Our primary dataset is the Medicare Current Beneficiary Survey (MCBS), an ongoing survey of the Medicare population (CMS, 199-2012). We use data from the population aged 65 and older, since Medicare is nearly universal in this age group. While data for the whole population would be preferred, data for the non-elderly population are less extensive, so we start with the elderly. This encompasses about one-third of national medical spending.¹⁴

The MCBS samples roughly 13,000 seniors annually. The survey gathers data on all use of medical care—both Medicare and non-Medicare. We modify MCBS data in two ways, discussed in more detail in Appendix A. First, Medicare claims data, and hence costs of care incurred, are only available for beneficiaries in traditional Medicare (i.e., not in Medicare Advantage HMO plans). To account for this, we reweight the population in traditional Medicare to reflect total Medicare enrollment, adjusting for the different demographics and health status of the two groups. Second, we adjust total spending in the MCBS by service category to match national estimates for the elderly (CMS, 2018). These adjustments are generally small and leave the trend in spending essentially unchanged (adjusted growth of 2.2% v. unadjusted growth of 2.1%).

Figure 2 shows per capita spending for elderly Medicare beneficiaries, in real 2010 dollars (adjusted using the GDP deflator). Spending rose from \$13,103 per capita in 1999 to \$17,897 per capita in 2012, an increase of \$4,794. There is a slowdown in spending growth around 2005, which has been examined (Cutler et al., 2019).

MCBS has a variety of sociodemographic and behavioral information, shown in Table 1. We present the means on these variables in five groups: sociodemographics (age, gender, race/ethnicity, marital status, poverty status, highest grade of education completed, and service in the armed forces), behavioral risk factors (smoking, body mass index [BMI]), self-reported health (one year prior and relative to others of the same age), mortality, and health impairments that we use to measure health-related quality of life, discussed in the next section.

A. Population Health

Our primary measure of health is quality-adjusted life expectancy, or QALE. At age a_0 ,

¹⁴ Trends in Medicare spending relative to costs would show up as increases or decreases in productivity. Hospital margins from Medicare are estimated to have declined in the few years after 1999 and been constant thereafter (MedPAC, 2018), but no analysis has looked at the trend in margins for all spending on the elderly.

QALE is defined as:¹⁵

$$(5) \quad \text{QALE}(a_0) = \sum_{t=0}^T S_t(a_0) \cdot H_t(a).$$

As noted above, we take a_0 to be age 65.

Survival rates are straightforward to measure, as age-specific mortality is tracked in Vital Statistics data (Centers for Disease Control and Prevention, 2002-2016).¹⁶ Figure 3(a) shows a rapid decline in the age-adjusted mortality rate in the elderly population from 1999 to 2012, 1.5% annually.

There is no universally accepted metric of health-related quality of life. Typically, quality of life is measured in domains, including physical and mental functioning, role limitations, pain, and cognition (Ware and Sherbourne, 1992). We term these domains ‘symptoms and impairments’ and denote the vector of prevalence rates as \mathbf{si}_t . \mathbf{h}_{si} is the corresponding vector of health decrements. Quality of life is then expressed empirically as $H_t(a) = \bar{H}_t(a) - \mathbf{si}_t(a) \cdot \mathbf{h}_{si}$, where $\bar{H}_t(a)$ is the upper limit of health, which may be less than 1 because of symptoms and impairments that are not included in our data. Both \mathbf{si}_t and \mathbf{h}_{si} may change over time, although for data reasons we hold the disutility of each symptom and impairment constant.

Our choice of symptoms and impairments, along with the health decrement weights, are determined by data availability (Stewart et al., 2013). The symptoms and impairments in MCBS are presented in the lower part of Table 1. A first metric is functional limitations, including difficulty walking, stooping/crouching/kneeling, lifting or carrying heavy objects, reaching or extending arms, and writing or grasping small objects. A second is self-care: whether respondents have impairments in any of 3 Activities of Daily Living (ADLs: dressing, bathing/showering, or walking) or 6 Instrumental Activities of Daily Living (IADLs: using the telephone, managing money, shopping, making meals, or doing light housework). We also use respondents’ reports of whether they have difficulty seeing or hearing, even when wearing glasses or a hearing aid. Finally, we use a report of whether health limits respondents’ social interactions, such as visiting with friends or close relatives. Appendix Figure A3 shows the trend in each of these measures of health over time, age-adjusted to the 2010 population. The changes in each are very small.

¹⁵ QALE is not discounted in this equation, whereas it is when we measure welfare. That is largely definitional. QALE is a physical concept. The value of those years is discounted.

¹⁶ Social Security Administration Life Tables (SSA, 2017) were used to calculate mortality rates for ages above the cutoff of 100 years in CDC life tables.

While this is a rich set of metrics, it is worth noting that several domains of health are not asked about explicitly in MCBS, including pain, cognition, and mental health symptoms. Appendix Figure A4 shows trends in these three measures from the Health and Retirement Study (HRS, 2021; RAND, 2021), age-adjusted to the 2010 population. The prevalence of mild and moderate pain increased over this time period. Cognition was flat, and depression fell modestly. To examine the potential impact of not having information on these dimensions, Appendix Figure A5 shows how pain, cognition, and depression are related to ADL impairments, IADL impairments, and functional limitations, which we have in our data. Higher levels of pain, cognition, and depression are associated with much higher probabilities of ADL and IADL limitations and more functional limitations. Thus, while we are missing some quality of life domains, we suspect the impact on our overall quality of life metric is not major.

To estimate the \mathbf{h}_{si} vector of health decrements for each symptom or impairment, we use data from the 2002 Medical Expenditure Panel Study (MEPS; AHRQ, 2002). This survey asks people questions about their health similar to the MCBS and also includes an overall health assessment on a scale from 0 ('worse imaginable health state') to 100 ('best imaginable health state').¹⁷ To form the disutility weight associated with each symptom and impairment, we regress the 100-point scale response on dummy variables for having each symptom or impairment.¹⁸ The results of the regression are shown in Appendix Table A5. The constant term, an estimate of \bar{H} , is 0.85, and is statistically different from 1. All of the coefficients for the symptoms and impairments are negative and statistically significant. The largest decrements are for health interfering with social interactions (-0.16 for moderate impairment and -0.20 for severe impairment) and for severe walking difficulty (-0.12).

We use these coefficients to calculate health-related quality of life for each person in each year of the MCBS.¹⁹ The minimum quality of life is 0.19; the average is 0.69. Figure 3(b) shows the trend in quality of life between 1999 and 2012. Consistent with the minor changes in the

¹⁷ The specific question is from the EQ-5D instrument: "To help people say how good or bad a health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked by 100 and the worst state you can imagine is marked by 0. We would like you to indicate on this scale how good or bad your own health is today, in your opinion. Please do this by drawing a line from the circle below to whichever point on the scale indicates how good or bad your current health state is."

¹⁸ We do not include demographic variables such as age because we do not want them to pick up the effect of symptoms and impairments. That said, the estimates associated with the symptoms and impairments do not differ greatly if demographics are included.

¹⁹ The decrements for symptoms and impairments are held constant over time, though the decrements for individual conditions change as prevalence changes.

prevalence of symptoms and impairments over time, quality of life was generally unchanged, with a very modest increase. Put another way; survival has increased without creating a larger impaired population.

The combination of falling mortality and relatively constant quality of life implies that quality-adjusted life expectancy increased. This is shown in Figure 3(c), which plots QALE at age 65 over time. In 1999, the typical person aged 65 could expect to live 12.1 additional quality-adjusted years. In 2012, the value was 13.0 years, a 1-year increase in just over a decade.

As discussed below, our benchmark value of a year of quality-adjusted life is \$100,000. Thus, the value of the QALE increase is roughly \$100,000. The increase in real, per capita medical spending, added up over remaining life, is about \$60,000 (\$4,800 per year for ~13 years). Thus, if the only thing that changed over time were medical treatments, the spending would be worth it. Of course, other factors changed as well. The goal of our productivity analysis is to estimate the value of medical treatment changes. Still, the results of the detailed empirical work are consistent with this first impression.

B. Identifying Conditions

There is an enormous number of medical conditions; over 68,000 are noted in the latest International Classification of Diseases (ICD, version 10). This is far too many for our analysis. The Agency for Health Care Research and Quality has consolidated this list into 285 Clinical Conditions (termed CCS; Elixhauser et al., 2014). However, even this condensed list is too granular for our purposes. For example, over 20 codes are for pregnancy and many others are for childhood conditions, each of which are uncommon in the elderly. In consultation with clinical experts, we reclassified the 285 CCS's into 78 conditions (Raghunathan et al., 2020), shown in Appendix Table A6.²⁰

One way to characterize these 78 conditions is by chapters of the ICD manual: circulatory disease, respiratory disease, and so on. We delineate highly prevalent conditions within each chapter and generally create an 'other' category for the remainder. For example, our cardiovascular disease conditions include ischemic heart disease, congestive heart failure, stroke and cerebrovascular disease, and 'other heart and vascular disease'—the latter of which is the residual.

²⁰ These conditions largely combine CCS categories, though some CCS categories are disaggregated given relatively high prevalence and differing disease courses in the elderly.

In some cases, there are no apparent subsets to pull out, and the entire chapter is left as one condition (as with gastrointestinal and liver disease). Some catchall categories are a bit broader. Two of our conditions are ‘after care’ (general sequelae to a hospitalization, for example physician visits for medication monitoring after surgery), and ‘signs and symptoms’ (adverse medical events of unknown origin, such as fainting or nausea). Seven of our categories are technically not conditions but preventive care: well care (such as medical checkups) immunization and screening for infectious disease, and cancer screening.

In addition to clinical conditions and preventive care, we create conditions for moderate and severe frailty. The frailty measures are based on self-reports of difficulty or inability to do 5 activities: walking 2-3 blocks, stooping/crouching/kneeling, lifting/carrying 10 lbs, reaching over one’s head, and writing or handling and grasping small objects. People with a lot of difficulty or an inability to do activities in 1-3 areas are classified as moderately frail, and people with difficulty in 4-5 areas are severely frail.

Most of the conditions that we measure have permanent impacts on health, for example heart disease and kidney failure.²¹ Thus, our measure of prevalence is whether the person ever had the condition. In principle, one might want to distinguish established cases from those that occurred more recently. However, the incidence of many conditions is too small to permit this.

Measuring condition prevalence accurately is important to our analysis. Given the observed improvement in aggregate health, increases in prevalence imply that the value of medical care is higher, since otherwise health would have fallen. To measure the prevalence of each condition, we start with diagnoses in the claims data in MCBS for all conditions except frailty, which is based on self-reports. Claims data are known to underreport prevalence, especially for less severe conditions. For example, 23% of people in the MCBS have a claim for arthritis in a year, but nearly half of the aged population in the National Health and Nutrition Examination Survey (NHANES) reports having been diagnosed with arthritis. The underdiagnosis changes over time; our data show declines in undiagnosed hypertension and high cholesterol over time.

To estimate true prevalence, we impute ‘calibrated conditions’ using self-reports of ever having had a condition in NHANES (CDC, 1999-2012). The procedure is described in detail in Appendix A and in Raghunathan et al. (2020). When a disease that has a self-report question in NHANES matches one of our categories (e.g., ischemic heart disease), we impute prevalence in

²¹ Exceptions can include injury and acute infectious disease that do not have long-term sequelae.

the MCBS to match the prevalence in NHANES, using demographic and health measures to match individuals across surveys. When there is not an exact self-report, we impute prevalence using conditions that have a similar overall prevalence rate to the relevant MCBS condition. In each case, we form five multiply-imputed datasets, each imputing different people to have each condition. Analyses using these data averages across the 5 data sets.

One important addition that comes from using the NHANES is that we can impute undiagnosed hypertension, high cholesterol, and diabetes, which are included in the NHANES from measurement and blood draws. As we show below, the treatment of cardiovascular risk factors is a significant part of health and spending changes for the elderly, so this imputation is quite valuable. Since the NHANES data are from 1999 onward, the beginning of our analysis period is limited to that year.

For presentation purposes, we generally aggregate the 80 conditions into 30 groups of relatively similar conditions. For example, acute myocardial infarction (heart attack) and coronary atherosclerosis (build-up of plaque in the coronary arteries) are grouped into ischemic heart disease. Within the 30 categories, trends tend to be relatively similar for the individual conditions. Appendix B shows all of our results for the 80 conditions.

The list of aggregated conditions, along with their prevalence in 2012, is shown in Figure 4. Cardiovascular risk factors are the most common condition; 86% of the elderly population has one or more. Other common conditions include musculoskeletal disorders (e.g., arthritis and back pain), eye and ear problems, respiratory problems, and uncategorized signs and symptoms. The presence of multiple chronic conditions is very common. The average elderly person has 13 conditions, and 95% of people have three or more conditions.

The prevalence of most conditions is rising over time (Appendix Table A6). The largest increases are for cardiovascular risk factors, chronic renal failure, and musculoskeletal issues, all of which are related to obesity. We return to the role of obesity in section IV. The major exception to the trend of rising prevalence is frailty. Moderate frailty was constant over time, and severe frailty fell.

III. Methodology for Allocating Spending and Health to Medical Care

Having assembled the data on health and medical spending, we need to estimate the impact of medical treatment changes on each. Figure 1 shows the analytic steps involved in this. We start

by estimating the direct effect of treatment changes on the health and medical spending of people with that condition. We then reallocate some spending and health outcomes from direct conditions to the exogenously identified medical risk factors. Finally, we estimate productivity change from trends in these estimates.

This analysis is particularly complex because there are many possible relationships across conditions. We need to pare this down to make progress. The primary assumption we make is that in most cases, medical care for a condition affects the health of people with that condition but not the prevalence of other conditions. That is, treatment changes for people with a heart attack influence the length and quality of life for people with a heart attack, but not whether the person develops cancer. The exception to this rule is conditions explicitly identified as risk factors, which we discuss below.

A. Estimating the Direct Effect of Each Condition

The first analytic step is to apportion total medical spending and health decrements to each condition. Recall from Section II that we have estimates of average medical spending ($m_{ave,t}$), mortality (Γ_t), and quality of life (H_t). These are the prevalence-weighted averages of the values for each condition:²² $m_{ave,t} = \mathbf{prev}_t \cdot \mathbf{m}_t$; $\Gamma_t = \mathbf{prev}_t \cdot \boldsymbol{\gamma}_t$; and $H_t = \bar{H}_t - \mathbf{prev}_t \cdot \mathbf{h}_t$. The issue is to estimate \mathbf{m}_t , $\boldsymbol{\gamma}_t$, and \mathbf{h}_t .

Aggregate medical spending for the elderly is estimated in the National Health Expenditure Accounts (CMS, 2021). Vital Statistics collects information on condition-specific mortality rates, for example death from ischemic heart disease (IHD) and lung cancer. These are per capita totals, i.e. $prev_{IHD_t} \cdot \gamma_{IHD_t}$. We could estimate the case fatality rate by dividing the per capita rate by prevalence. However, vital statistics reported underlying cause of death data are notoriously inaccurate. Vital Statistics attribute each death to one primary or underlying cause of death, but many deaths have multiple antecedents, as our example of mental illness suggests. The physician or coroner certifying death may not know about precursor events, and even if they do, they may not record them.²³ This is the analog to the issue of condition enumeration noted above; physicians tend to record the severe event but less commonly the comorbid conditions that contributed to it.

²² For notational convenience, we omit the age subscripts.

²³ Because death is determined when the heart stops, heart disease is overcoded on death records relative to other causes (Messite and Stellman, 1996).

We develop a methodology to parcel overall health and medical spending to conditions that accounts for these issues (Ghosh et al., 2020, 2021). The methodology is described briefly here and in greater detail in Appendix B. The analysis relies on propensity score models. For each year, we estimate a logistic model for the probability of having each condition—80 regressions in total. The covariates include the other 79 conditions along with the demographic and behavioral risk factor variables shown in Table 1, including smoking status and obesity.²⁴

Based on the propensity scores, we stratify the population into quintiles, or quartiles/terciles for conditions with relatively low prevalence.²⁵ The difference in the within-strata outcomes of people with and without the relevant condition, averaged across strata and across the five multiply-imputed datasets, is an estimate of relative mortality associated with each condition. Term this relative mortality $\hat{\theta}_{ct}^{\gamma}$, where γ denotes mortality, c is the condition, and t is the year. The analogues for quality of life and medical spending are $\hat{\theta}_{ct}^h$ and $\hat{\theta}_{ct}^m$. The total mortality rate is allocated to conditions based on these relative mortality rates:²⁶

$$(6) \quad \hat{\gamma}_{ct} = \Gamma_t \cdot \left(\frac{\hat{\theta}_{ct}^{\gamma}}{\sum_c \hat{\theta}_{ct}^{\gamma}} \right),$$

The analog of equation (6) for spending is somewhat more complex than for mortality because the distribution of spending is so skewed to high spenders. High spenders do not have different conditions than low or moderate spenders; rather, their high costs reflect common conditions whose course of therapy goes very poorly. As a result, when the condition-specific costs are then added up to the person level, they come nowhere close to reproducing the skewed distribution of individual spending (Appendix Figure B6). We thus concluded that a uniform adjustment to spending would not be appropriate.

To better capture the variability of spending, we use an alternative to equation (6). We relate the actual cost for each person, m_{it} , to the predicted cost based on summing the first stage estimates to the individual level, $\hat{\theta}_{it}^m$, interacted with a series of variables capturing high use of

²⁴ Including the other conditions in the model implies independence conditional on the other X's. In cases where there is clear causal relationship with other conditions, as with risk factors and their associated direct conditions, we model the link directly. In other cases where there is a deterministic or extremely tight correlation with the condition of the interest, we exclude or sample on certain covariates. For example, undiagnosed hypertension is not included in the model for diagnosed hypertension, and gender is not included in the model for prostate cancer. Appendix B provides more detail.

²⁵ We use this method as opposed to a 1:N match or weighting because the sample of people with many conditions is small. See Appendix B.

²⁶ For quality of life, $\sum_c \hat{\theta}_{ct}^h$ is very close to the population estimate, H_t . Thus, we set $\hat{h}_{ct} = \hat{\theta}_{ct}^h$.

services (\mathbf{X}_{it}):

$$(7) \quad m_{it} = \hat{\theta}_{it}^m \cdot (\mathbf{X}_{it} \boldsymbol{\alpha}_t) + \varepsilon_{it}.$$

\mathbf{X}_{it} includes the number of conditions and its square, an indicator for any hospitalization, the number of admissions, and the number of nights in the hospital, the number of days institutionalized, survival in the year, the number of months survived for those who are deceased, and the number of outpatient claims. We average the $\mathbf{X}_{it} \hat{\boldsymbol{\alpha}}_t$ estimates across people with each condition to form a condition-specific correction factor, $\hat{\varphi}_{ct}$, which is then multiplied by the first stage cost estimates: $\hat{m}_{ct} = \hat{\theta}_{ct}^m \cdot \hat{\varphi}_{ct}$. This effectively increments the spending of conditions where people have very extensive care utilization. Appendix Table B2 shows the estimates of equation (7) in 2009. The \mathbf{X} variables lead to increased spending in the expected direction; for example, people who spend more time in a hospital or an institution cost more than people who do not.

Figure 5 shows an example of the resulting spending and health outcomes for coronary atherosclerosis and heart attack. Together, these make up the category of ischemic heart disease, which is among the categories where we find the highest productivity growth (as discussed below). While there are some outlier years, the trend is generally smooth, with declining mortality and relatively constant cost.

The validity of the conditional unconfoundedness assumption cannot be tested completely. One empirical test is to examine covariate balance for people with and without each condition. Table 2 shows the cohort balance for coronary atherosclerosis and acute myocardial infarction in 1999. Panel A shows the balance for variables included in the model. The overall F-statistic fails to reject equality of the means between cases and controls, and only 2 of the 40 comparisons are statistically different. As shown in Appendix Figure B3, we fail to reject equality of the means between cases and controls for all conditions in all years (N=1120).

We also examined the balance of six covariates which were excluded from the model because they were judged not likely to be related to spending. These include whether care is regularly received in a doctor's office or elsewhere, the source of Medicare supplemental insurance if any, out-of-pocket spending, uncollected liability, and residence in a rural location. Panel B of Table 2 shows no significant differences between cases and controls in these excluded variables. Appendix Figure B4 shows that we reject equality of the means between cases and controls for only 3% of the observations. Using data in 1999, Appendix Figure B5 shows that the standardized

mean difference for each of the omitted conditions is very small. Thus, we do not find evidence of confounding from those variables we have chosen to omit.

Spending and health decrement can be read directly from the values for each year. However, because the sample sizes of people with each condition are not always large, each year is measured with some error. To minimize this error, we fit a second-order polynomial in year for each of the annual estimates of mortality, quality of life, and medical spending. The fitted values for the first and last year are then used to estimate productivity growth. The fitted values for heart disease are shown in Figure 5.

B. Accounting for Clinical Risk Factors

Our second analytic step is to reallocate some of the spending and health outcomes from direct conditions to clinical risk factors. Conceptually, the idea is that some of the direct effect estimated for conditions such as ischemic heart disease is better seen as attributable to hypertension or high cholesterol, which led to the heart disease in the first place.

To determine which conditions are risk factors and what direct conditions they contribute to, we consulted with a physician working group and examined published information. Figure 6 shows the resulting risk factors and their influenced direct conditions. The set of risk factors is reasonably conservative, to avoid inappropriately taking spending and health decrements away from direct conditions. Of our 80 conditions, four are classified as exclusively risk factors: high cholesterol, hypertension, eye and ear disorders, and infectious disease immunization (flu and pneumococcal vaccine). Two conditions, diabetes and mood disorders, are both direct conditions and risk factors for other diseases.²⁷

To determine the amounts to be reallocated, we estimate regression models relating each direct condition to its identified risk factors. The results of these regressions are shown in Appendix Table B5. A natural way to express the results is as relative risks—the probability of having the direct condition if one has the risk factor divided by the probability if one does not. Where possible, we compare our relative risks to those in the clinical literature; they match up reasonably well.

To demonstrate the resulting adjustment, we use the example of hypertension leading to

²⁷ Certain other conditions are highly correlated. For example, people who are frail suffer more falls. There is debate as to whether this is causal or whether it instead reflects common conditioning factors. In the absence of general agreement in the clinical literature, we use a limited set of risk factors.

heart disease. Pooling 1999-2001, our regression estimate is that people with hypertension are 7.5 percentage points more likely to have heart disease than are people without high blood pressure. At that time, 57% of people had high blood pressure and 35% of people had ischemic heart disease. This implies that 12% of cases of ischemic heart disease are due to hypertension ($7.5\% \times 57\% / 35\%$). Thus, we reduce the prevalence of heart disease by 12%. We make an offsetting entry in the hypertension industry. People with heart disease spend on average \$1,100 per person more than similar people without heart disease. We therefore transfer \$1,100 per transferred case to the hypertension industry. This amounts to \$47 per person with hypertension ($\$1,100 \times 7.5\% \times 57\%$). This gets added to the direct spending on hypertension, estimated as above. Appendix Table B6 shows the full set of risk factor adjustments.

We make the same conceptual adjustment with mortality and quality of life, transferring decrements associated with direct conditions to risk factors. In this case, conditions that are purely risk factors are assumed to have no direct effect on death or quality of life, so the entirety of their impact on health is through the risk factor channel.

In addition to the risk factors, four of our conditions are cancer screening variables: breast, colorectal, prostate, and cervical cancer. Conceptually, these are closely related to risk factors, in that they influence health by affecting direct conditions. The difference is that cancer screening does not reduce the incidence of cancer in total. Rather, it reduces metastatic cancer, but increases local cancer. Indeed, the increase in local cancer is generally larger than the reduction in metastatic cancer, as some cases are detected that would never become metastatic (Welch and Browley, 2018).²⁸

If we had data on cancer by stage, we could separately model the impact of cancer screening on local and metastatic disease. However, our data do not contain information on stage. As a result, we use impacts of cancer screening drawn from the clinical literature. As shown in Appendix Table B5, the clinical literature shows clear benefits of breast and colorectal screening in reducing mortality, but unclear benefits of cervical and prostate cancer screening, especially in the elderly. We thus model only mortality benefits from breast and colorectal cancer screening. We assume no impact of cancer screening on quality of life, since quality of life likely varies with treatment, and treated cases rise with screening. Because of these approximations, we note that the

²⁸ Colon cancer screening through colonoscopy is a hybrid; it both detects cancer earlier and reduces the development of colon cancer, for example by removing polyps.

estimate of productivity growth estimate for cancer screening is subject to more uncertainty than are other conditions.

C. Productivity Growth Estimates

Our third analytic step is to estimate productivity growth, the net value of the change in QALE less the increase in medical spending (all in real 2010 dollars). The change in QALE is found by simulating the impact of changes in each condition individually on length and quality of life. These are then combined in equation (5). Spending is found by adding condition-specific spending in the relevant year over the remaining lifetime. To value QALE in dollars, we use a baseline value for a year of life in good health equal to \$100,000 (Viscusi and Aldy, 2003). We conduct sensitivity analyses assuming the value is alternatively twice and half that amount. We use a baseline discount rate of 0% and perform sensitivity analyses using a rate of 3%.

To express productivity as a growth rate, we need to divide by the net value of medical care in the base year. As a rough estimate, we assume that half of QALE in 1999 is due to medical care (Cutler, Rosen, and Vijan, 2006).

D. Standard Errors

We use a bootstrap technique to calculate standard errors for the productivity growth estimates. We draw 1,000 simulated samples, taken from the actual MCBS population with replacement. We do this separately for each multiply-imputed dataset. For each draw, we repeat the three steps above. These estimates are used to calculate the empirical 95% confidence interval for the spending, health, and productivity growth impacts of each condition.

E. Model Uncertainty

In addition to classical uncertainty, it is worth noting that we do not have random variation in treatments with which to measure the impact of medical care changes. As a result, our evaluation of health and spending changes is a form of ‘residual analysis’—the difference for people with and without each condition after controlling for other factors that influence health.

To validate our most important results, we undertake several supplemental analyses, shown in Appendix C. These include examining how our results change if we use Vital Statistics estimates of changes in mortality by condition rather than our estimated rates; omitting quality of life;

estimating condition-specific spending using different methodologies; and smoothing the time series in other ways. None of these changes has a large impact on our findings.

We also compare our findings about sources of health improvement to microsimulation estimates of the impact of treatment taken from validated and widely used clinical models. We focus particularly on cardiovascular disease and cancer, the source of our most important results. As we show below, our results match the clinical literature very closely.

IV. Change in Medical Spending by Condition

In this section, we present our estimates of spending by condition. The first two columns of Table 3 show spending per case for the 30 aggregated conditions in 1999 and 2012 (all in 2010 dollars). Appendix Table B4 shows direct spending for each of the 80 conditions. Almost all conditions are associated with higher spending. The major outlier to this is cancer screening, where the estimates suggest slightly lower spending for people who are screened relative to those who are not. It may be true that cancer screening saves money, though it is also possible that elderly people who are screened for cancer are healthier in ways that our model does not capture. In some studies (Trogdon, 2008), condition-specific spending is constrained to be non-negative. That may not be necessary for our application, however, as we are interested in the change in spending at the condition level. If the bias from unmeasured health status is constant over time, the *change* in spending will be accurate even if the *level* of spending is not. Thus, we do not constrain spending to be positive. That said, caution is warranted about the estimated cost changes for cancer screening.

The conditions for which per case spending is highest are generally the ones we would expect: heart disease, certain cancers (especially lung cancer), dementia, kidney failure, and frailty. The case-specific costs may seem low, but it is important to recall that these are the cost per prevalent case of disease, not per incident case. For example, spending on heart attacks is the average among everyone who ever had a heart attack, not new heart attack cases. Because past cases generally add a smaller increment to spending than new events, our estimates are smaller than are typically found in studies looking at acute disease onset.

Spending for most conditions rose over time. The largest increases were for lung cancer (a \$2,500 per case increase) and arthritis and musculoskeletal problems (an \$750 per case increase).

Notably, the conditions that cost the most at a point in time are not always the ones for which spending rose the most. For example, spending for ischemic heart disease rose only modestly over the time period, despite starting at a high level.

Figure 7 shows the change in per capita spending—per case spending times the prevalence rate—accounted for by each of our 30 aggregated conditions. Acute cardiovascular diseases together account for 15% of spending in 1999, but only 0.5% of the spending increase. The largest contributors to increased spending are cardiovascular risk factor conditions, musculoskeletal diseases (which includes arthritis and back pain), and after care. Together, these three conditions account for 40% of the increase in spending.

Spending per capita can increase because the prevalence of disease increases or because each case costs more. The importance of each of these is shown in the first column of Table 4. Sixty percent of the increase in real spending is due to the greater cost per case; one-third is due to greater prevalence (the residual is the covariance term). Prevalence changes were particularly significant for cardiovascular risk factors, renal failure, and musculoskeletal problems. These three conditions are among those whose prevalence is most affected by obesity. To determine how much of the prevalence increase is due to obesity, we use the same methodology as for clinical risk factors. We estimate the impact of BMI and smoking on the prevalence of conditions that are listed in clinical reviews as resulting from these risk factors.²⁹ We then simulate the impact of changes in smoking and obesity on the prevalence of both direct conditions and risk factors.

Table 4 shows the simulated impact of smoking and obesity changes on overall medical spending via their impact on disease prevalence. Increased obesity led to an increase in spending of \$342 per person. Smoking declines were more modest, a reduction of \$64 per capita. There is a sizeable residual increase in the prevalence of many conditions not associated with smoking or obesity changes, \$1,382, shown in the last row of the table.³⁰

²⁹ Appendix Table B9 compares the values we estimate to those in the clinical literature. Behavioral risk factors can also influence spending conditional on disease, but we have held these constant in our methodology.

³⁰ It is possible that the impact of obesity on spending is understated, as we only have data on current weight and conditions are likely to be affected by weight in the past in addition to current weight. While we attributed a reduction of 0.15 years of life expectancy to current BMI, Preston, et al. (2018) attributed a reduction of 0.9 years life expectancy to lifetime maximum BMI over almost the same time period (1998-2011) among those 40 and older. Since is the nearly same amount of life expectancy loss that we attributed to ‘other factors’ affecting disease prevalence, it could be that obesity accounts for essentially all of the disease prevalence change. It is also possible that environmental changes such as increases in stress have led to greater prevalence of disease.

V. Change in Health by Condition

In this section, we decompose changes in health into the same conditions for which we estimate the change in spending. Estimates of mortality and quality of life impact for our 30 aggregated conditions are shown in the third–sixth columns of Table 3; results for all 80 conditions are in Appendix Tables B7 and B8. The direct conditions are almost all associated with higher mortality and lower quality of life than not having the condition. Further, the relative magnitude makes sense; people with lung cancer have an annual mortality rate 15% greater than comparable controls (this includes past as well as current cases of lung cancer); acute kidney failure adds a 3% increment to mortality.

The mortality and quality of life data allow us to estimate QALE for people with each condition in 1999 and 2012, and thus the change in QALE. These results are shown in Figure 8. The largest increase in QALE was for cardiovascular diseases—particularly ischemic heart disease but including other cardiovascular diseases as well. People with cardiovascular disease experienced an increase in QALE of 1 year over similar people without cardiovascular disease. Many other conditions had increases in QALE that were somewhat smaller than this but still substantial. For example, people with lung cancer, colorectal cancer, and acute kidney failure all experienced QALE increases of around 0.4 years. Frail people lived an additional 0.3 quality-adjusted years over the time period. Most of the increase was in life expectancy, but quality of life improved for people with cardiovascular disease and some other conditions.

Table 4 shows the combined impact of medical treatments and non-medical changes in prevalence on QALE at age 65. The overall increase in QALE noted above (1.04 years) is the net effect of two offsetting trends. Holding prevalence constant, QALE would have increased by 1.69 years. However, the rising prevalence of conditions reduced QALE by -0.86 years. Some of this reduction in QALE was a result of increasing obesity (-0.16 years). This effect, while only a portion of the prevalence increase, is quantitatively very large. For example, the negative impact of obesity on QALE is roughly equivalent to the entire QALE increase from improved care of musculoskeletal conditions and mental health conditions combined.

VI. Productivity Growth in Medical Care

Having estimated the cost and health consequences of changes in medical treatments, this section brings the two together to measure productivity growth. Figure 9 and Table 5 show the

calculation of productivity change by condition and for the medical system as a whole.

The first column in Table 5 shows the increase in the present value of lifetime costs, the capitalized version of the changes in Table 3. The second column is the increase in QALE. The remaining columns show productivity growth estimates for different values of V (the value of a year of life) and ρ (the discount rate). Standard errors are presented in Table 5 for the overall productivity growth estimate and in Appendix Table B11 for each of the conditions.

Several results are apparent. The overall estimate of medical care productivity growth, shown in the first row of Table 5, is positive. On net, medical care increased in value by \$113,000. This estimate is statistically significantly different from zero. The value of health improvement is roughly \$170,000, offset by an increase in spending of \$58,000. To express this as a growth rate, we need to divide by the net value of medical care in the base year. Assuming that half of QALE at age 65 in the base year is due to medical care, we estimate that medical care productivity growth was 21% in aggregate, or 1.5% per year.³¹

The high overall productivity growth estimate is a mix of heterogeneous estimates for different conditions, shown visually in Figure 9. Cardiovascular disease had the highest productivity growth rate. The net benefit from cardiovascular disease was \$89,000 per case. Other conditions have more modest but still large increases in value-added. For example, the net benefit from treatment changes is \$38,000 for people with colorectal cancer, \$30,000 for people with acute kidney failure, and \$22,000 for frail people.

Very few conditions have negative productivity growth, but many are near 0. For example, there was little productivity change for people with arthritis and musculoskeletal conditions (-\$3,000 on net) or mental illness (\$4,000 on net), even though treatment changed rapidly for both conditions. The declining productivity of treating musculoskeletal conditions likely reflects the increased use of opioids over this period. These drugs are expensive but do not appear to be very effective in reducing chronic pain (Phillips, Ford, and Bonnie, 2017). The low productivity change of mental illness is more difficult to understand. Despite a vast increase in the number of people treated with drugs for mental illness, life expectancy and quality of life among those with mental health diagnoses varied little. In interpreting both of these findings, it is worth reiterating that we do not have direct measures of pain and depressed mood in our data. Still, the measures we have

³¹ In 1999, medical spending was about \$13,000 per year and life expectancy was 12.2 years. Assuming half of this is due to medical care, the health benefit would be 6.1 years, at a cost of roughly \$80,000.

are highly correlated with them.

Columns 4 and 5 of Table 5 show the impact of varying the value of a life-year, to \$50,000 in column 4 and \$200,000 in column 5. These changes affect the net value of medical care as one would expect. Medical care as a whole has a value of about \$25,000 at the lower value, with a 95% confidence interval that includes 0. The higher value of life raises the net benefit to just above \$280,000.

The final column of Table 5 shows the impact of using a 3% discount rate for both cost and utility discounting. The results show the same pattern, though the overall magnitude is somewhat smaller. The net benefit of medical care is \$50,000 per person.

A. Sensitivity Analysis

We have explored the sensitivity of these results in several ways, detailed in Appendix C. First, we examined robustness to the measure of mortality change by condition. As expected, our methodology attributes far fewer deaths to heart disease in each year than does Vital Statistics—half as many to ischemic heart disease. However, our estimates of *changes* in death by cause are similar to Vital Statistics. Thus, productivity growth for cardiovascular disease is similar using the Vital Statistics data as our estimates: \$77,000 using Vital Statistics data vs. \$89,000 using the propensity score method.

Second, we consider productivity estimates using only length of life, since mortality is easier to measure than quality of life. As Appendix C shows, results using length of life only are very similar to those in Figure 9.

Third, we consider how our estimates of spending by condition compare to those using other methodologies. We examine two alternative methodologies: one that parcels out spending for each claim to conditions listed on that claim, and a second that uses a single regression model to parcel out total spending during the year to conditions coded during the year (see Appendix C for details). Using data from 2009, we show that spending by condition is reasonably highly correlated using the three methods: correlations are generally on the order of 0.7 to 0.8. Thus, while we have not estimated costs using these three methods in each year, we suspect that the impact of using alternative spending methodologies would be modest.

Finally, we compare results from our smoothed estimates of health and spending changes to estimates using averages of the first and last few years of the data. The change in health and

medical spending is very similar using both methodologies.

B. Cardiovascular disease

Because cardiovascular disease is so prominent as a cause of death, researchers have developed clinical models of cardiovascular disease mortality reductions to which we can compare our results. The most commonly referenced heart disease model is the IMPACT model (Ford et al., 2007; Ogata et al., 2019). The IMPACT model sorts people into cells based on the presence or absence of heart disease and uses clinical trial evidence on the impact of different treatments to simulate how treatment changes have affected mortality. For example, one cell of the model estimates the differential mortality rate for people with prior heart disease who are and are not taking antihypertensive medications. Using this coefficient and the share of people taking antihypertensive medications over time, the model estimates the change in mortality that is predicted to have occurred because of the expanded treatment of hypertension. We use the parameters of the IMPACT model to benchmark our findings.³²

In addition to looking at heart disease, we replicated the structure of the IMPACT model for congestive heart failure and cerebrovascular disease, the other major forms of cardiovascular disease. In some cases, the studies that provide the impact of treatments on mortality are the same as those for ischemic heart disease. In other cases, we find comparable meta-analyses.

Table 6 shows the comparison of our results to the IMPACT model. The first column shows national totals for cardiovascular disease mortality. Relative to a constant age and sex-specific death rate over time, cardiovascular disease mortality in the elderly declined by 300,000 deaths between 1999 and 2012. The results of the IMPACT model are shown in the second column. The model suggests there should be 219,000 fewer deaths resulting from medical treatments over the time period.

The estimates from our model, presented in the third column, are that medical care accounts for 233,000 fewer deaths over the time period. This is within 17% of the IMPACT model estimates. One would expect our results to be somewhat greater than the IMPACT model since we allow for all effects of cardiovascular risk factors, which include non-cardiovascular complications such as kidney disease. Thus, the estimates are very close.

³² Details are in Appendix D. The impact of treatments on outcomes is generally not estimated differently for the elderly and non-elderly populations.

In addition to matching model totals, our results add richness to them. In the national data, very few deaths are attributable to cardiovascular risk factors, and in the IMPACT model, deaths due to risk factor management are not separated out. Our data shows that a good deal of the improvement in cardiovascular disease mortality comes from better control of risk factor conditions. That is, the prevention industry is important alongside the treatment industry.

C. Cancer

We can also compare our results for certain cancers to those in the clinical literature. Table 5 shows large productivity gains in lung, colorectal, and ‘other’ cancers (most commonly skin and cervical cancer). QALE for each of these conditions increased about 0.4 years. We use data from the National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) program (NCI, 1999-2012) to benchmark the lung and colorectal cancer results.³³

The trend in mortality that we attribute to lung cancer matches well with national data. SEER data indicate that mortality per case fell 10% between 1999 and 2012; our estimate is a reduction of 8%. For colorectal cancer, there is a larger discrepancy: an 8% reduction in per case mortality in SEER, compared to 35% in our data. It is not clear what explains this difference, but it suggests that our estimate of productivity growth in colorectal cancer may be too high.

The clinical literature helps evaluate whether the survival gains for these conditions are likely due to medical treatment advances. The major alternative explanation for improved survival is that survival gains are an artifact of detecting smaller cancers that would never become fatal or detecting cancers earlier in the process, so that survival appears to improve.³⁴ Bias from increased diagnosis is unlikely to be an issue in lung cancer, where screening was minimal over the time period and most tumors grow rapidly (Noone et al., 2017). Further, treatments have improved over time. Lung surgery is generally less invasive, and chemotherapy is more common, as are biomarker-targeted therapies. Van der Drift et al. (2012) and Gunn et al. (2018) argue that these treatment changes have improved lung cancer survival in the Netherlands and Finland.

³³ The SEER data give survival in each year after diagnosis for people diagnosed in each calendar year. To calculate the mortality rate in a calendar year, for example 1999, we find mortality rates in the first year among people diagnosed in 1999, in the second year among people diagnosed in 1998, and so on. We weight these mortality rates by the share of people alive in the relevant year who would have been diagnosed in each past year.

³⁴ This goes by several names, including lead time bias (the treatment has no impact but the cancer is known about for longer), length bias (screening detects more slowly growing cancers so that the average mortality rate among clinically treated cancers falls) and overdiagnosis (the detection of preclinical cancers that would never have become clinically relevant).

Screening for colorectal cancer grew over the time period, therefore diagnosis changes could be an explanation for improved survival. However, other evidence suggests that medical technology changes have materially improved survival for colorectal cancer. Treatment changes include greater use of pre-surgical radiation, increased use of adjuvant chemotherapy, new medications for metastatic disease, and development of new targeted biologic agents. Colorectal cancer models suggest that greater use of therapies known about in 2000 could lead to reduced mortality on the order of what occurred (Vogelaar et al., 2006). Lemmens et al. (2010) and Brouwer et al. (2018) argue that these treatment changes can explain the improvement in case-specific colorectal cancer survival that was observed in the Netherlands. However, there is no definitive evidence for the United States as yet.

VII. Conclusion

Measuring productivity growth in medical care is a central policy concern. In this paper, we model the productivity of the medical sector as a whole along with specific conditions. We do so in the context of a satellite health account, an account designed to complement current accounts, which focuses largely on medical spending.

Our analysis is detailed, but the facts underlying the conclusions are straightforward. First, there has been a large increase in survival for the elderly. The survival increase occurred despite a rise in obesity and thus condition prevalence. Using conventional estimates of the value of life, the gain in QALE attributable to treatment changes is valued much more than the increase in spending. The net is a productivity increase of 21% in total, or 1.5% annually.

Second, the most important contributor to rising productivity is cardiovascular disease. Cardiovascular risk factors are better controlled over time, and interventions for acute disease have improved survival. The improved health we find for people with cardiovascular disease is also found in aggregate evidence and is consistent with clinical trial data on the impact of particular therapies. These gains have been achieved at very modest cost increase, leading to very high productivity growth.

Third, medical spending increases were spread across many conditions, including many with little health improvement. The prototypical example is musculoskeletal disease, where pharmaceutical use and surgical interventions grew rapidly but overall health was unchanged. The same is true for mental health, though this is subject to more uncertainty. As a result,

musculoskeletal and mental health conditions are estimated to have low productivity growth.

There are several estimates of health care productivity in the economic literature (Sheiner and Malinovskaya, 2018). However, comparing our estimates with others is difficult. Conceptually, some studies look at changes in the average cost of a year of life over time (e.g., Romley et al., 2015, 2020). An increased average cost per year of QALE is judged to reflect declining productivity. In contrast, we look at the marginal value of health changes relative to their marginal cost. The idea that increasing average cost reflects low productivity growth is true only if one could have bought additional years of quality-adjusted life at the same price in the base year. In that case, spending more per QALY over time is a productivity reduction. If technical constraints prohibited buying more health in the base period, however, an increasing average cost of QALE improvement is not a sign of declining productivity. Rather, one needs to evaluate the marginal cost and benefit of treatment changes. Comparing across studies is also difficult because different studies use different measures of quality. Quality metrics in past studies include clinical estimates of the value of innovation (Lakdawalla et al., 2015), predicted survival based on biomarkers (Eggleston et al., 2011), and expert opinion about the value of treatment changes (Berndt et al., 2002). It is unclear how these relate to actual changes in outcomes. Even studies that look at health outcomes (Romley et al., 2015, 2020) do so over different time horizons.

The most comparable papers to ours are Cutler et al. (1998) and Dauda et al. (2019). Cutler et al. estimate price reductions of 14% annually for elderly heart attack sufferers. Dauda et al. estimate price reductions of 7% annually for elderly people hospitalized with heart attack, heart failure, or pneumonia conditions. The price reductions in both of these studies are larger than we estimate for the average elderly person, reflecting the very high benefits of treatment change for acute cardiovascular events.

Our analysis has several limitations, which would be valuable to address in future research. We have estimated the value of medical care for the elderly. However, Case and Deaton (2017) show different trends in mortality in the adult non-elderly populations. Similarly, our findings do not provide guidance on which types of care have been most and least valuable over this time period. Expanding the health account to cover such topics is a worthy area for future research.

REFERENCES

- Abraham, Katherine G., and Christopher Mackie, eds., *Beyond the Market: Designing Nonmarket Accounts for the United States*, Washington, D.C.: National Academy Press, 2005.
- Abraham, Katherine G., and Christopher Mackie, “A Framework for Non-Market Accounts,” in Dale Jorgenson, J. Steven Landefeld, and William D. Nordhaus, eds., *A New Architecture for the U.S. National Accounts*, Chicago: University of Chicago Press, 2006, 161–192.
- Advisory Commission to Study the Consumer Price Index. 1996. “Toward a More Accurate Measure of the Cost of Living.” Washington, DC: US Senate Finance Committee, 104th Congress, 2nd session, S Print 104-72.
- Agency for Healthcare Research and Quality (AHRQ). 2004. “MEPS 2002 Full Year Consolidated Data File.”
https://meps.ahrq.gov/mepsweb/data_stats/download_data_files_detail.jsp?cboPufNumber=HC-070 (accessed August 27, 2021).
- Berndt, Ernst R., David M. Cutler, Richard Frank, Zvi Griliches, Joseph P. Newhouse, and Jack E. Triplett, “Medical Care Prices and Output,” in Anthony Culyer and Joseph Newhouse, eds., *Handbook of Health Economics, Volume I*, Amsterdam: Elsevier, 2000, 119–180.
- Berndt, Ernst R., Anupa Bir, Susan H. Busch, Richard G. Frank, and Sharon-Lise T. Normand, “The Medical Treatment of Depression, 1991-1996: Productive Inefficiency, Expected Outcome Variations, and Price Indexes,” *Journal of Health Economics*, 2002, 21(3): 373–396.
- Brouwer, Nelleke P.M., Amand C.R.K. Bos, Valery E.P.P. Lemmens, Pieter J. Tanis, Niek Hugen, Iris D. Nagtegaal, Johannes H.W. de Wilt, and Rob H.A. Verhoeven, “An Overview of 25 Years of Incidence, Treatment, and Outcome of Colorectal Cancer Patients,” *International Journal of Cancer*, 2018, 143(11): 2758–2766.
- Brynjolfsson, Erik, Avinash Collis, W. Erwin Diewert, Felix Eggers, and Kevin J. Fox, “GDP–B: Accounting for the Value of New and Free Goods in the Digital Economy,” NBER Working Paper No. 25695, March 2019.
- Bureau of Labor Statistics, “Technical Information about the US Multifactor Productivity Measures,” mimeo, September 26, 2007.
- Case, Anne, and Angus Deaton. “Mortality and Morbidity in the 21st Century,” *Brookings Papers on Economic Activity*, Spring 2017.
- Centers for Disease Control and Prevention. 1999-2012. “National Health and Nutrition Examination Survey” years 1999-2012.
<https://www.cdc.gov/nchs/nhanes> (accessed August 27, 2021).

- Centers for Disease Control and Prevention (CDC). 2002-2016. “National Vital Statistics report United States Life Tables,” years 1999-2012.
https://www.cdc.gov/nchs/products/life_tables.htm#life (accessed August 27, 2021).
- Centers for Disease Control and Prevention, National Center for Health Statistics. 1999-2012. “Underlying Cause of Death 1999-2019 on CDC WONDER Online Database, released in 2020. Data are from the Multiple Cause of Death Files, 1999-2019, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program.” <http://wonder.cdc.gov/ucd-icd10.html> (accessed Sept 1, 2021).
- Centers for Medicare & Medicaid Services. 1999-2012. “Medicare Current Beneficiary Survey (MCBS),” years 1999-2012. <https://www.cms.gov/Research-Statistics-Data-and-systems/Research/MCBS> (accessed August 27, 2021).
- Centers for Medicare and Medicaid Services. “National Health Expenditure Accounts,” years 1999-2012. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData> (accessed June 9, 2021).
- Chandra, Amitabh, and Jonathan Skinner, “Technology Growth and Expenditure Growth in Health Care,” *Journal of Economic Literature*, September 2012, 50(3): 645–80.
- Cockburn, Iain, and Aslam H. Anis, “Hedonic Analysis of Arthritis Drugs,” in David M. Cutler and Ernst Berndt, eds., *Medical Care Output and Productivity*, Chicago: University of Chicago Press, 2001.
- Comanor, William S., Stuart O. Schweitzer, and Tanja Carter, “A Hedonic Model of Pricing of Innovative Pharmaceuticals,” in Marco R. Di Tommaso and Stuart O. Schweitzer, eds., *Health Policy and High-Tech Industrial Development*, Edward Elgar Publishing, 2005.
- Cutler, David M., Mark McClellan, Joseph P. Newhouse, and Dahlia Remler, “Are Medical Prices Declining? Evidence from Heart Attack Treatments,” *The Quarterly Journal of Economics*, 113(4), November 1998, 991–1024.
- Cutler, David M., Allison B. Rosen, and Sandeep Vijan, “Value of Medical Spending in the United States: 1960–2000,” *New England Journal of Medicine*, 2006, 355(9): 920–927.
- Cutler, David M., Kaushik Ghosh, Kassandra Messer, Trivellore Raghunathan, Susan T. Stewart, and Allison B. Rosen. “Explaining the Slowdown in Medical Spending Growth among the Elderly, 1999–2012,” *Health Affairs*, 38(2), February 2019, 222–229.
- Dauda, Seidu, Abe Dunn, and Anne Hall, “Are Medical Care Prices Still Declining? A Systematic Examination of Quality-Adjusted Price Index Alternatives for Medical Care,” mimeo, 2019.
- Dunn, Abe, Lindsey Rittmueller, and Bryn Whitmire. 2015, “Introducing the New BEA Health Care Satellite Account,” *Survey of Current Business*, 2015, 95(1), 1–21.

- Eggleston, Karen N., Nilay D. Shah, Steven A. Smith, Amy E. Wagie, Arthur R. Williams, Jerome H. Grossman, Ernst R. Berndt, Kirsten H. Long, Ritesh Banerjee, and Joseph P. Newhouse. "The Net Value of Health Care for Patients With Type 2 Diabetes, 1997 to 2005," *Annals of Internal Medicine*, 2009, 151(6): 386–393.
- Elixhauser Anne, Claudia Steiner, and L. Palmer. Clinical Classifications Software (CCS). U.S. Agency for Healthcare Research and Quality, <https://www.hcup-us.ahrq.gov/toolssoftware/ccs/CCSUsersGuide.pdf> Internet accessed, 2018.
- Ford, Earl S., Umed A. Ajani, Janet B. Croft, Julia A. Critchley, Darwin R. Labarthe, Thomas E. Kottke, et al., "Explaining the Decrease in U.S. deaths from Coronary Disease, 1980–2000," *New England Journal of Medicine*, 2007, 356(23): 2388–2398.
- Fraumeni, Barbara M., Michael S. Christian, and Jon D. Samuels, "The Accumulation of Human and Nonhuman Capital, Revisited," *Review of Income and Wealth*, 2017, 63 (s2): S381–S410.
- Ghosh, Kaushik, Irina Bondarenko, Kassandra L. Messer, Susan T. Stewart, Trivellore Raghunathan, Allison B. Rosen, and David M. Cutler. "Attributing Medical Spending to Conditions: A Comparison of Methods," *PLoS One*, 2020, 15(8):e0237082.
- Ghosh, Kaushik, David Cutler, Trivellore Raghunathan, Kassandra Messer, Susan T. Stewart, and Allison B. Rosen. "Attributing Mortality to Medical Conditions: Levels and Trends in the United States, 1999–2012," mimeo, 2021.
- Groshen, Erica L., Brian C. Moyer, Ana M. Aizcorbe, Ralph Bradley, and David M. Friedman, "How Government Statistics Adjust for Potential Biases from Quality Change and New Goods in an Age of Digital Technologies: A View from the Trenches," *Journal of Economic Perspectives*, 2017, 31(2): 187–210.
- Gunn, Jarmo, Johanna Valo, Jussi Sipilä, Päivi Rautava, Eero Sihvo, and Ville Kytö, "Trends and Results of Lung Cancer Surgery in Finland between 2004 and 2014," *European Journal of Cardio-Thoracic Surgery*, July 2018, 54 (1): 127–133.
- Health and Retirement Study (HRS) public use datasets, 1998-2012. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI. <https://hrs.isr.umich.edu/> (accessed August 27, 2021).
- Jones, Charles I., and Peter J. Klenow. 2016. "Beyond GDP? Welfare across Countries and Time." *American Economic Review*, 106 (9): 2426-57.
- Kuznets, Simon, *National Income 1929–1932*. Senate Document No. 124, 73rd Congress, 2nd Session. Washington, D.C.

- Lakdawalla, Darius, Jason Shafrin, Claudio Lucarelli, Sean Nicholson, Zeba M. Khan, and Tomas J. Philipson, "Quality-Adjusted Cost of Care: A Meaningful Way to Measure Growth in Innovation Cost Versus the Value of Health Gains," *Health Affairs*, 2015, 34(4): 555–61.
- Landefeld, J. Steven, and Stephanie H. McCulla. "Accounting for Nonmarket Household Production within a National Accounts Framework," *Review of Income and Wealth*, September 2000, 46(3): 289–307.
- Lemmens Valery, Liza van Steenberg, Maryska Janssen-Heijnen, Hendrik Martijn, Harm Rutten, and Jan Willem Coebergh. "Trends in colorectal cancer in the south of The Netherlands 1975-2007: rectal cancer survival levels with colon cancer survival," *Acta Oncology*, 2010, 49:784–96.
- Medicare Payment Advisory Commission, *Medicare Payment Policy: Report to the Congress*, Washington, D.C.: MedPAC, March 2018.
- Melnikow, J., Jillian T. Henderson, Brittany U. Burda, Caitlyn A. Senger, Shauna Durbin, and Meghan S. Weyrich. "Screening for Cervical Cancer with High-Risk Human Papillomavirus Testing: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force," *JAMA*, 2018, 320(7):687–705.
- Messite, Jacqueline, and Steven D. Stellman, "Accuracy of Death Certificate Completion: The Need for Formalized Physician Training," *JAMA*, 1996, 275(10):794–6.
- Muller, Nicholas Z., Robert Mendelsohn, and William Nordhaus, "Environmental Accounting for Pollution in the United States Economy." *American Economic Review*, 2011, 101(5): 1649–75.
- National Cancer Institute (NCI), Surveillance, Epidemiology, and End Results (SEER) Program. 1999-2012. "SEER*Stat Database: Mortality - All COD, Aggregated With State, Total U.S. (1969-2018) <Katrina/Rita Population Adjustment>," DCCPS, Surveillance Research Program, released May 2020. <http://www.seer.cancer.gov> (accessed August 30, 2021). Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
- National Research Council, Panel to Advance a Research Program on the Design of National Health Accounts. Washington (DC): National Academies Press (US); 2010.
- Noone, Anne-Michelle, Kathleen A. Cronin, Sean F. Altekruse, Nadia Howlader, Denise R. Lewis et al., "Cancer Incidence and Survival Trends by Subtype Using Data from the Surveillance Epidemiology and End Results Program, 1992–2013," *Cancer Epidemiology, Biomarkers, and Prevention*, 2017, 26(4):632–641.
- Nordhaus, William D. "The Health of Nations: The Contribution of Improved Health to Living Standards." In *Measuring the Gains from Medical Research: An Economic Approach*, edited by Kevin M. Murphy and Robert Topel, 9–40. Chicago: University of Chicago Press, 2013.

- Ogata Soshiro, Kunihiro Nishimura, Maria Guzman-Castillo, Yoko Sumita, Michikazu Nakai, Yoko M. Nakao, Nobuo Nishi, et al., “Explaining the Decline in Coronary Heart Disease Mortality Rates in Japan: Contributions of Changes in Risk Factors and Evidence-Based Treatments between 1980 and 2012,” *International Journal of Cardiology*, 2019, 291(September):183–188.
- Phillips Jonathan K., Morgan A. Ford, and Richard J. Bonnie, eds., *Committee on Pain Management and Regulatory Strategies to Address Prescription Opioid Abuse*, National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Health Sciences Policy; Washington (DC): National Academies Press (US); 2017 Jul 13.
- Preston, Samuel H., Yana C. Vierboom, and Andrew Stokes, “The Role of Obesity in Exceptionally Slow US Mortality Improvement,” *Proceedings of the National Academy of Sciences*, 2018, 115, 201716802.
- Raghunathan, Trivellore E., Kaushik Ghosh, Allison B. Rosen, Paul Imbriano, Susan T. Stewart, Irina Bondarenko, Kasey Messer, Patricia Berglund, James Shaffer, and David M. Cutler, “Combining Information from Multiple Data Sources to Assess Population Health,” *Journal of Survey Statistics and Methodology*, March 2020, smz047.
- RAND HRS Longitudinal File 2018 (Feb 2021 (V1) and RAND HRS Fat Files, 1998-2012. Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration. Santa Monica, CA. <https://www.rand.org/well-being/social-and-behavioral-policy/centers/aging/dataproduct.html> , <https://hrsdata.isr.umich.edu/data-products/rand> (accessed August 27, 2021).
- Romley, John A., Dana P. Goldman, and Neeraj Sood, “US Hospitals Experienced Substantial Productivity Growth During 2002–11,” *Health Affairs*, 2015, 34(3): 10–1377.
- Romley, John A., Dana P. Goldman, Neeraj Sood, and Abe Dunn, “Productivity Growth in Treating a Major Chronic Health Condition,” mimeo, Brookings Institution, 2020.
- Rosen, Allison B., and David M. Cutler. “Measuring Medical Care Productivity: A Proposal for U.S. National Health Accounts,” *Survey of Current Business*, June 2007, 87: 54–58.
- Rosen, Allison B., and David M. Cutler. “Challenges in Building Disease-Based National Health Accounts.” *Medical Care* 47(7)_Supplement_1 (July 2009): S7–13.
- Rosen, Allison B., Eli Liebman, Ana M. Aizcorbe, and David M. Cutler. “Comparing Commercial Systems for Characterizing Episodes of Care,” Bureau of Economic Analysis (BEA) Working Paper WP2012–7. 2012, Washington, DC: BEA.
- Schultze, Charles and Christopher Mackie, eds. *At What Price?: Conceptualizing and Measuring*

- Cost-of-Living and Price Indexes*. Washington, D.C.: National Academies Press, 2002.
- Sensenig, Arthur, and Ernest Wilcox, “National Health Accounts/National Income and Product Accounts Reconciliation—Hospital Care and Physician Services,” in David Cutler and Ernst Berndt, eds., *Medical Care Output and Productivity*, Chicago: University of Chicago Press, 2001, 271–302.
- Sheiner, Louise, and Anna Malinovskaya, “Measuring Productivity in Health Care: An Analysis of the Literature,” Brookings Institution, May 2016.
- Social Security Administration (SSA), 2017. “Period Life Tables.”
<https://www.ssa.gov/oact/HistEst/PerLifeTables/2017/PerLifeTables2017.html> (accessed August 27, 2021)
- Stewart, Susan T., David M. Cutler, and Allison B. Rosen. 2013. “US trends in quality-adjusted life expectancy from 1987 to 2008: combining national surveys to more broadly track the health of the nation.” *American Journal of Public Health* 103, no. 11 (November 2013): e78–87.
- Trogdon, Justin, Eric A. Finkelstein, and Thomas J. Hoerger, “Use of Econometric Models to Estimate Expenditure Shares,” *Health Services Research*, August 2008, 43(4): 1442–1452.
- Van der Drift, Miep A., Henrike E. Karim-Kos, Sabine Siesling, Harry J.M. Groen, Michel W.J.M. Wouters, Jan-Willem Coebergh, Esther de Vries, and Maryska L.G. Janssen-Heijnen, “Progress in Standard of Care Therapy and Modest Survival Benefits in the Treatment of Non-Small Cell Lung Cancer Patients in the Netherlands in the Last 20 Years,” *Journal of Thoracic Oncology*, February 2012, 7(2): 291–298.
- Vogelaar, Iris, Marjolein van Ballegooijen, Deborah Schrag, Rob Boer, Sidney J. Winawer, J. Dik F. Habbema, and Ann G. Zauber, “How Much Can Current Interventions Reduce Colorectal Cancer Mortality in the U.S.? Mortality Projections for Scenarios of Risk-Factor Modification, Screening, and Treatment,” *Cancer*, 2006, 107(7): 1624–1633.
- Viscusi, W. Kip, and Joseph E. Aldy, “Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World,” *Journal of Risk and Uncertainty*, 2003, 27(1), 5–76.
- Ware, John E., and Cathy D. Sherbourne, “The MOS 36-Item Short Form Health Survey (SF36). I. Conceptual Framework and Item Selection,” *Medical Care*, June 1992, 30(6): 473-483.
- Welch, H Gilbert and Otis W. Brawley, “Scrutiny-Dependent Cancer and Self-fulfilling Risk Factors,” *Annals of Internal Medicine*, 2018, 168(2):143–144.