Health Spending Growth Has Slowed: Will the Bend in the Curve Continue?

Joseph P. Newhouse
Harvard University

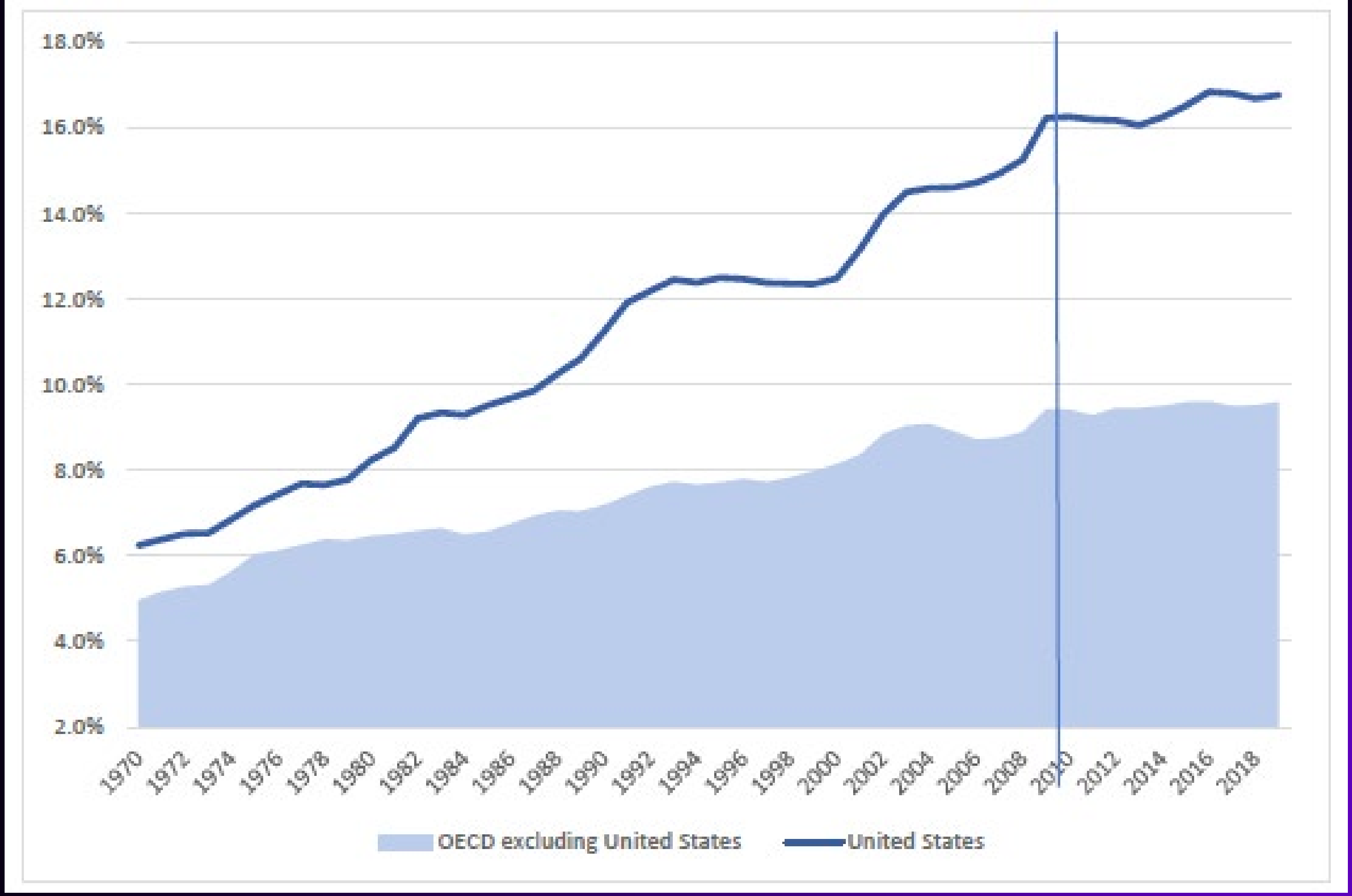
My Coauthors

- Are Sheila Smith and Gigi Cuckler from the Office of the Actuary, Center for Medicare & Medicaid Services, US Department of Health and Human Services (CMS)

Nothing Grows to the Sky

- As the next slide shows, health spending as a share of GDP grew rapidly after 1970 in both the US and many other OECD countries, but slowed dramatically after 2009
  - The OECD data come from the 19 OECD countries for which there are data for the 1970-2019 period; those countries are:
    - Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Japan, Korea, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, and the UK
  - We stopped the time series at 2019 because of the pandemic

Health Spending as a Share of GDP in the US and the OECD Excluding the US*



*Here and elsewhere "OECD" refers to the 19 countries shown on the prior slide.

The Figure on the Prior Slide Motivates the Two Questions We Seek to Answer

- What caused the growth in health care spending as a share of GDP to greatly slow starting in 2009 in the US and the 19 other high-income countries?
- What does the answer to that question imply about future health care spending growth?
  - That last question is not just of academic interest; the Office of the Actuary at CMS, where my two co-authors work, is charged by law with projecting future health care spending growth*

*Their projections appear annually in the Medicare Trustees' Report and can be found on the CMS website, https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsProjected. CBO also projects future health spending.

We Build on the Shoulders of a Giant

- We borrow from Robert Solow's classic 1957 paper that found technical change was the major source of growth in the overall US economy*
- Under some assumptions Solow showed that

$$\frac{\dot{Q}}{Q} = \frac{\dot{A}}{A} + w_K \frac{\dot{K}}{K} + w_L \frac{\dot{L}}{L}$$

where the dotted terms in *Q*, *K*, and *L* are growth rates of output, capital, and labor, respectively. The dotted term *A*, a residual, is the proportion of growth Solow attributed to technological change. The w are factor shares.

*Robert Solow, "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, August 1957.

Solow's Conclusion

- Although couched with numerous (and justified) caveats, Solow concluded that the great majority, around 7/8, of the increase in output per labor hour in the US economy between 1909-1949 was attributable to technical change

In 1992 I Borrowed from Solow's Method to Look at Health Care Spending Growth*

- I tried to account for how much growth in health care spending above overall economic growth from 1940-1990 could be attributed to 4 factors: *demographics*, primarily aging; the spread of *insurance*; *income* growth; and below average *productivity* growth in a service industry
- Like Solow's conclusion for the entire economy, I concluded these 4 factors could not explain the majority of the spending increase and that the main driver of the increase was the increased capabilities of medicine, or technical change

**Journal of Economic Perspectives*, Summer 1992.

Smith, Newhouse, and Freeland, 2009

- In 2009 Sheila Smith, Mark Freeland, and I took another run at accounting for health care spending growth*
- Analogously to Solow and my 1992 paper, our 2009 paper used the following equation:

$$\dot{H}/H = \dot{A}/A + \varepsilon_y(\dot{Y}/Y) + \varepsilon_i(\dot{I}/I) + \varepsilon_d(\dot{D}/D) + \varepsilon_p(\dot{P}/P)$$ where

$\dot{H}/H$, $\dot{Y}/Y$, $\dot{I}/I$, $\dot{D}/D$, and $\dot{P}/P$ are growth rates of health care spending, income, insurance, demographics, and excess relative price growth, respectively, and $\dot{A}/A$ is a residual

*Smith, Newhouse, and Freeland, *Health Affairs*, September 2009.

Smith, Newhouse, and Freeland, cont.

- We, like Solow and my 1992 paper, assumed $\dot{A}/A$, the residual, was predominantly attributable to technological change in medicine
- The ε's we used came from the literature; we tested the sensitivity to assuming income elasticities of 0.6 and 0.9 and annual productivity change in medical care between 0 and the economy-wide average of 0.8%
  - We needed the assumption on productivity change in medicine to derive $\dot{P}/P$, or excess relative price growth

Conclusions from the 2009 Paper

- We estimated that technical change accounted for between 27% and 48% of the growth in US health care spending between 1960 and 2007
- While still substantial, that was less than my 1992 estimate because I had used an income elasticity at the household level, whereas the 2009 paper's assumed income elasticities were based on the larger values from the literature that used GDP to explain health care spending at the area or country level; as a result, income accounted for more of the growth

Exogenous and Endogenous Technical Change

- Our 2009 paper also made an advance over my 1992 paper by decomposing the residual, or technological change, into exogenous and endogenous change
- Exogenous change is change that would occur even if income (or the resource constraint) is held constant; endogenous change is induced by growth in income

Teasing Apart Endogenous and Exogenous Change

- Econometrically, endogenous change can be represented as an interaction term between income and technical change
- Technical change in medicine is common across high-income countries; we assumed that year fixed effects, holding income constant, measured exogenous technical change. This is a strong assumption: it implies either new technology is introduced everywhere simultaneously or new technology appears at a constant rate over time and each country's lag structure in introducing it is stable

That Brings Me to Our Current Work

- Sheila Smith, Gigi Cuckler, and I are now updating the 2009 paper, using not only 1970-2019 data from the US but also from the 19 other OECD countries ("OECD exUS") for which there are data
- I am mostly going to skip the details of the econometrics and focus on the results; we can certainly come back to the details if you like,* but I do need to mention our treatment of income (real GDP per capita) and medical price inflation

*More details are available in our National Bureau of Economic Research Working Paper #30782.

At the Country Level Changes in GDP Have Both Immediate and Lagged Effects

- When household incomes or consumer prices change, a household can make many adjustments quickly, e.g., buy less orange juice if its price rises
- But when GDP changes, the prices households pay for health care don't necessarily change right away; e.g., deductibles and copays may stay fixed for a period of time, so responses are not immediate

Changes in GDP, cont.

- Similarly, prices providers receive for their services don’t change quickly when GDP changes; at the public level such changes will typically require legislation; at the private level insurers may have multi-year contracts with providers
- Furthermore, when prices do change providers may not respond immediately

So Income for Our Purposes Includes Both Current and Past Income

- We include both contemporaneous and lagged income to explain variation in health spending as a % of GDP; we use real GDP per capita as the income variable
- We tested for a lag in the effect of changes in GDP of up to 8 years, but the best fit to the data was a moving average of GDP in the current year and the prior 4 years with equal weights on each year*

*We specified a polynomial lag to allow for unequal effects by year but did not find them.

Relative Unit Medical Prices

- Spending = *p* x *q*, but we only have medical price data (*p*) for the US and even those data are only reliable after 1995
  - For the US between 1970-1995 we use BLS data on medical sector input prices and high and low assumptions on medical sector total factor productivity relative to economywide productivity
- For the OECD we have no data and so omit the relative price term; we also estimate a comparable equation for the US, which we term the "Simplified Decomposition"

The Results for 1970-2019: Income and Technology Drive Spending Growth*

Table 2: Share of Growth in Real per Capita Spending on Health Consumption Expenditures Attributed to Causal Factors, 1970–2019

1970 to 2019					
	United States			Simplified decomposition[2]	
	Estimated range[1]		Midpoint	US	OECD exUS
(a) Income effect	36.6%	41.5%	39.0%	39.0%	52.2%
(b) Demographic change[3]	8.5%	8.5%	8.5%	8.5%	12.9%
(c) Insurance coverage[4]	6.2%	6.2%	6.2%	6.2%	2.1%
(d) Relative medical price[4]	5.6%	10.3%	7.9%		
(e) Technology [=100-(a+b+c+d)]	43.1%	33.6%	38.3%		
(f) Income-Technology Interaction	24.3%	21.2%	22.7%	22.7%	30.4%
(g) US Technology Residual [=100-(a+b+c+d+f)]	18.8%	12.4%	15.6%		
(h) OECD exUS residual [= 100-(a+b+c+f)]				23.5%	2.4%

The "Simplified" decomposition omits the relative price term, so both it and the exogenous technology effects are in the residual. This allows us to show comparable numbers for the US and the OECD exUS.

*The "estimated range" reflects low and high assumptions on medical price inflation for the 1970-1995 period, the period before US Producer Price Indices were available. The simplified decomposition: both exogenous technology and the relative price term are in line h.

But for 2009-2019 the Picture Differs

2009 to 2019					
	United States			Simplified decomposition[2]	
	Estimated range[1]		Midpoint	US	OECD exUS
(a) Income effect	39.2%	44.3%	41.8%	41.8%	61.6%
(b) Demographic change[3]	29.1%	29.1%	29.1%	29.1%	35.1%
(c) Insurance coverage	6.0%	6.0%	6.0%	6.0%	0.5%
(d) Relative medical price	-6.6%	-6.6%	-6.6%		
(e) Technology [=100-(a+b+c+d)]	32.3%	27.1%	29.7%		
(f) Income-Technology Interaction	25.9%	22.7%	24.3%	24.3%	35.9%
(g) US Technology Residual =100-(a+b+c+d+f)	6.3%	4.4%	5.4%		
(h) OECD exUS residual = 100-(a+b+c+f)				-1.2%	-33.1%

Four Differences in the 2009-2019 Subperiod vs 1970-2019 Are Noteworthy

- *Demographics* accounts for a much larger share of the (small) growth in GDP share;
  - In the US: 29.1% 2009-2019 vs. 8.5% 1970-2019; in the OECD: 35.1% vs. 12.9% 1970-2019
  - This reflects baby boomer aging and the smaller total growth
- In the US the effect of *relative price inflation* changes sign, from +7.9% to -6.6%; we do not have a comparable measure for the OECD exUS

Four Differences, cont.

- In the 2009-2019 period the share of growth attributable to *technology overall* continues to be large in the US, 29.7%, but falls to near zero in the OECD exUS, 2.8% (=35.9-33.1)
- But the *technology residual* (exogenous technology) declines markedly for both the US and the OECD exUS, 16% to 5% in the US and 2.4% to -33%(!) in the OECD exUS*

*The OECD residual includes both the technology residual and the relative price effect, but any plausible value for the relative price effect in the OECD exUS would still show a large decline in the technology residual.

And One Important Inference

- In both the US and the OECD the share of growth attributable to income change not only remained large but even rose, in the US from 39% 1970-2019 to 42% in 2009-2019 and in the OECD from 52% to 62%
- The large (and even increased) shares of growth attributable to income together with the 5-year lag on the effect of income mean the Great Recession was an important factor in the slowdown in both the US and the OECD

What Does All This Imply for the Future?

- First a caveat: The following comments abstract from Covid-19, whose effect on spending we assume (and hope!) will become negligible in a few more years
- *Demographics*: Although the US is now past the peak contribution of aging to health spending, the effect of aging on spending will continue to remain well above the historical mean for the next two decades before tapering off*

*See the 2023 Medicare Trustees Report, page 66.

What Does This Imply for the Future?, cont.

- Given a 5-year lag on the *income* effect, we are now well past the effects of the Great Recession; unless the Fed's recent interest rate increases induce more than a modest recession, income effects going forward should approximate steady-state effects
- We also expect *relative medical unit price inflation* (relative to economywide inflation) to have a negligible effect on health's GDP share

Changes in Exogenous Technology

- Relative to the entire 1970-2019 period, there were striking declines in the portion of growth attributable to exogenous technology in 2009-2019 in both the US (15.6% vs 5.4%) and the OECD (2.4% vs -33.1%)
  - Unlike the US figures, the OECD figures include a relative price effect in both periods, but even if that effect isn't approximately the same in the OECD in the two periods (and so difference out), it isn't plausible that it could account for very much, if any, of a 35 percentage point decline

Interpreting the ΔGrowth Attributable to ΔExogenous Technology

- Exogenous technical change is a mix of innovation that is either:
  - Costly but sufficiently beneficial that it would be introduced even if income were constant; or
  - Saves money (assuming sufficiently little offset in quality)

Interpreting the ΔGrowth Attributable to ΔExogenous Technology, cont.

- Hence the fall in the proportion of growth due to exogenous technology is some combination of:
  - A lower proportion of highly beneficial innovation;
  - A higher proportion of cost saving innovation
- Can we find some "corroborative detail, intended to give artistic verisimilitude to an otherwise bald and unconvincing narrative"?*

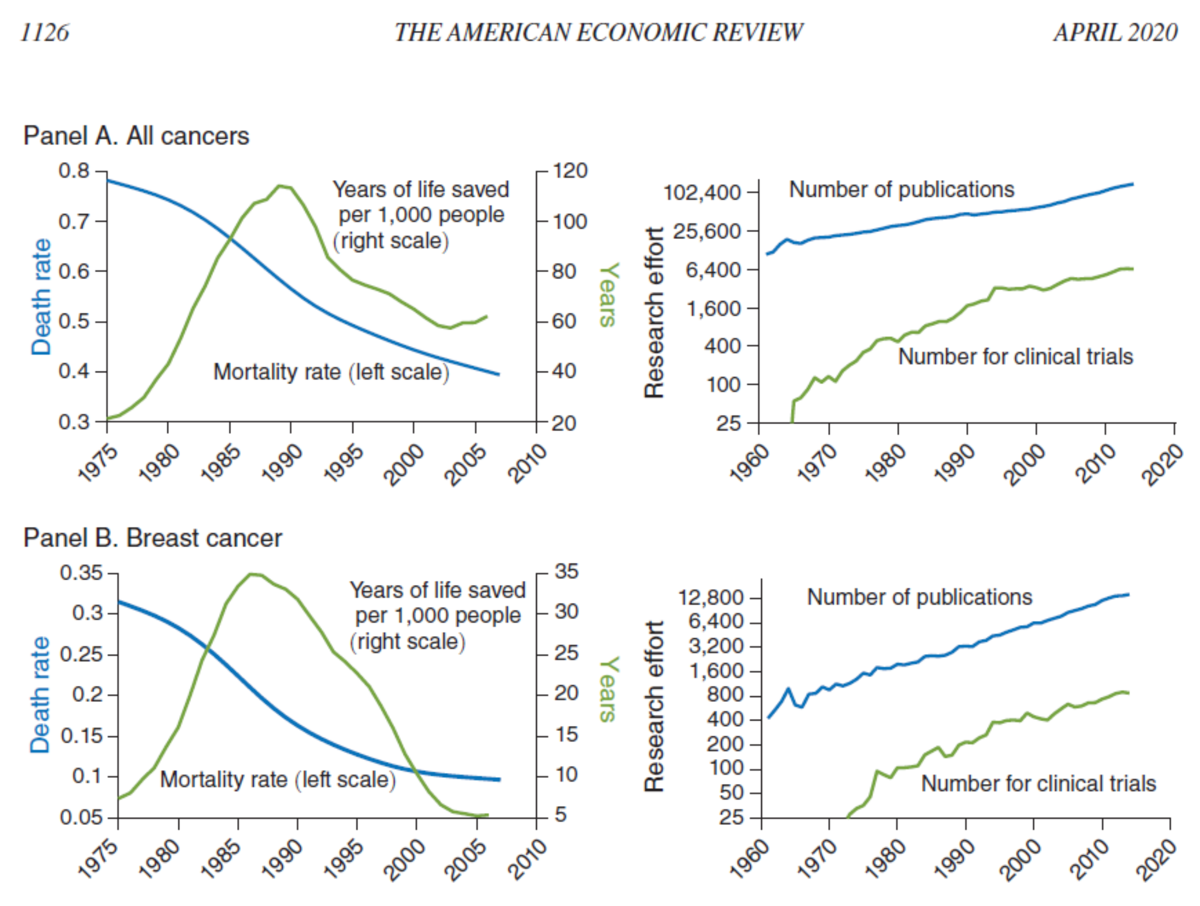
**From W.S. Gilbert's lyrics for *The Mikado* (a line spoken by Pooh-Bah).

A Diminishing Returns Story: "Ideas" Are Getting More Expensive

- Bloom, et al. show that in many fields measures of research output per researcher are falling*
  - For example, Moore's Law says the number of transistors on a chip doubles every 2 years, but it now takes about 18 times as many researchers to accomplish that as it did 50 years ago
- On the next slide Bloom, et al. show something similar for biomedical research; if beneficial innovation is getting more expensive, it is plausible that its rate of growth could fall

**American Economic Review*, April 2020, 110(4):1104-44.

More Research Effort Generates Smaller Life Expectancy Gains



The green lines in the left two panels are years of life expectancy gained annually for all cancers and for breast cancer. After 1990 both green lines fall despite the ever increasing research inputs shown in the right two panels. Bloom, et al. show a similar result for heart disease.

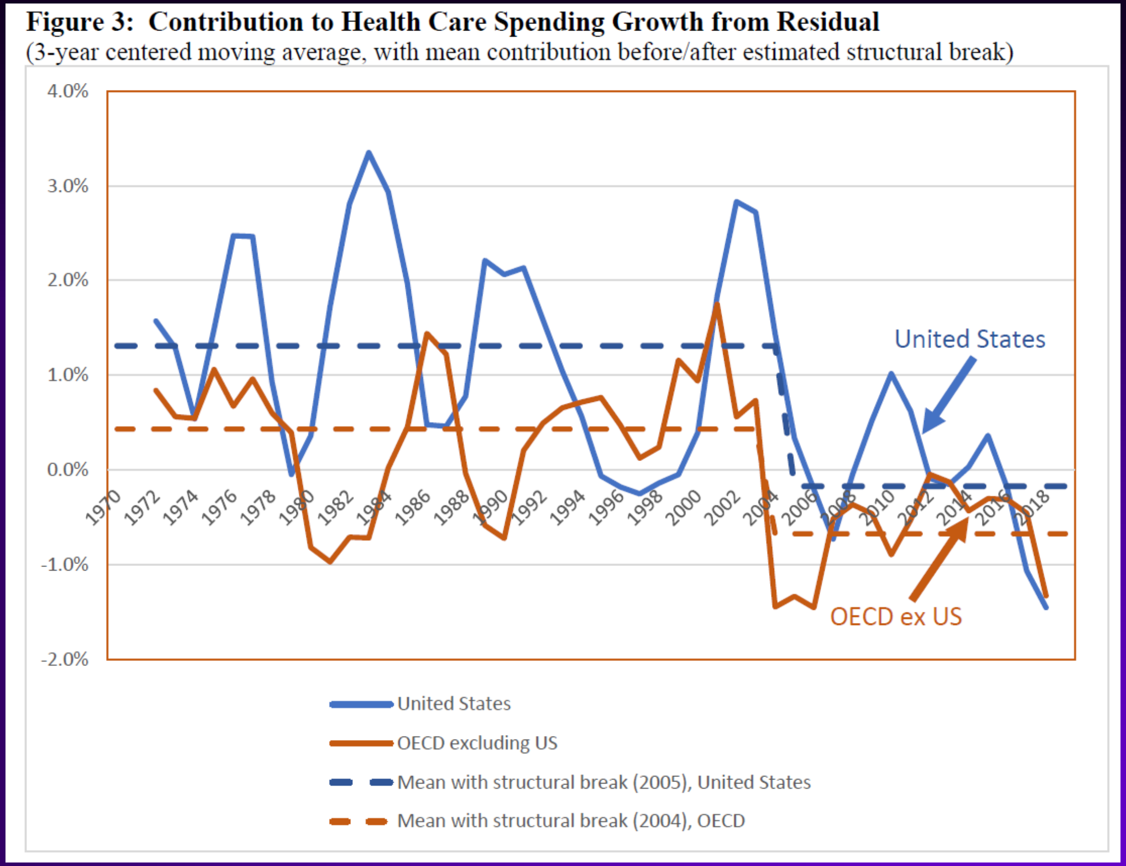
There Is a Literature on Induced Innovation*

- As health care's share of GDP has risen, one would expect demand for cost saving innovation would rise
- Further, the exogenous technology residual fell in both the US and the OECD; other than demographics and income (the Great Recession), both of which are controlled for, the only other major factor that would seem to be shared between the US and the rest of the OECD is technology

*Acemoglu and Linn, *Quarterly Journal of Economics*, August 2004, 119(3):1049-90, Agha et al., *American Economic Review: Insights*, June 2022, 4(2):191-208, Blume-Kohout and Sood, *Journal of Public Economics*, January 2013, 97:327-36, Finkelstein, *Quarterly Journal of Economics*, May 2004, 119(2):527-64.

We Tested for a Structural Break in the Residual*

We rejected the null of no structural break at p=0.0002 for the US and p=0.0000 for the OECD; the most likely date was 2005 for the US and 2004 for the OECD. This implies that exogenous technology had a smaller cost increasing effect after 2004-2005. This is earlier than the 2009 break that we used above; we chose 2009 based on the figure above on raw spending, not the residual.



*We used a Quandt-Andrews Test

Further Evidence

- We also thought this “story” would be more convincing if it were widely shared across countries, so we estimated the sign and size a of discrete country-specific change in the residual in 2004 for the entire period 1970-2019; the sign was negative in 17 of the 20 countries
  - We could reject the null at 5% in 14 of the 17 and in only one country, South Korea, was the sign significant and positive; South Korea is unusual because of its large decline in out-of-pocket share, 50% to 32% 1997-2019 vs 20% to 17% for all the OECD*

*The large change in out-of-pocket share suggests there may have been other changes to insurance that made it more generous in ways that out-of-pocket share would not control for.

Exogenous Technology and the Future

- Assuming the effect of exogenous technology remains at its lower post-2004-2005 level, looking forward we expect less of a positive contribution to spending growth from technology than in the 1970-2019 period and certainly less than in the 1970-2004 period
- Indeed, if technology innovation were to focus even more on cost saving, its contribution to spending growth in the US could continue to exert downward pressure – as appears to have happened in the rest of the OECD

Summing Up

- The slow growth in health care spending as a % of GDP 2009-2019 was due to lagged effects of the Great Recession, a fall in the role of exogenous technological change, and in the US a fall in relative prices, offset by baby boomer aging
- Looking forward, aging will push up health care spending for the next two decades in both the US and the OECD, but the fall in the role of exogenous technological change should lead to slower rates of growth than was the case, 1970-2019

And It's Always a Good Idea to Keep Yogi Berra's Advice in Mind

When forecasting Yogi counseled humility: "It's hard to make predictions, especially about the future."

Writing Out the Equations We Used

$$(2.1)\quad \ln(H_t) - \beta_I \ln(I_t) - D_t = \alpha + \beta_y \mathrm{MA}(\ln(Y_t), 5) + \sum_{c=0}^{I} c_i + \sum_{t=0}^{T} yr_t + \mu_{it}$$

$$(2.2)\quad \ln(H_t) - \beta_I \ln(I_t) - D_t = \alpha + \beta_y' \mathrm{MA}(\ln(Y_t), 5) + \sum_{c=0}^{I} c_i + \mu_{it}$$

H_t	= real per capita spending on health consumption, time t
I_t	= out-of-pocket share of health care spending, time t
D_t	= demographic index, time t
Y_t	= real per capita GDP, time t
c_i	= fixed effects for each country i
yr_t	= fixed effects for each year t
$\mathrm{MA}(\ln(Y_t), 5)$	= moving average of $\ln(Y_t)$, 0 to 5 years

We assume the year-fixed effects account for exogenous change in equation 2.1 and endogenous change is in the residual μ, whereas both types of change are in the residual in equation 2.2.

Estimating Exogenous Technological Change

- Because exogenous change is in the residual of equation 2.2 but not in the residual of 2.1, we can use the omitted variable theorem and the estimated difference in income elasticities, $\beta' - \beta$, to distinguish the two types of change*
- To estimate equations 2.1 and 2.2 we used 1970-2019 panel data from the US and the 19 OECD countries listed above
  - Although our 2009 paper used data from the 1960's, we now begin in 1970 because of data availability from a larger number of countries

*$\beta' - \beta$ is the coefficient of the interaction term between income and technological change assuming that the residual in equation 2.1 measures primarily endogenous technical change and the residual in equation 2.2 measures the combined effect of exogenous and endogenous change.

The Results from Estimating Equations 2.1 and 2.2

Step 1:	**Estimation of Equations 2.1 and 2.2 20-country sample[1], 1970-2019**				
	β_y		β'_y		$\beta'_y - \beta_y$
Estimated coefficient (standard error)[2]	0.8848 (0.0351)		1.3380 (0.0081)		0.4531 (0.0361)

β' exceeds β because it incorporates the income x technical change interaction.

Standard errors are clustered on country to account for serial correlation.

The Generosity of Insurance Variable *I*

- We measure *I* as the % paid out-of-pocket, or the average coinsurance rate; this is crude, but is the best that can be done with country as the unit of observation
- We did not estimate β_I, the coefficient of ln(*I*), but rather assumed it to be -0.1, the elasticity of spending with respect to the average coinsurance rate in the RAND Health Insurance Experiment*

*Manning, et al., *American Economic Review*, June 1987, Table 9.

The Demographic Variable *D*

- For the US *D* is mean spending for age-sex-time-to-death cells, weighted by the population share of the cell
  - Time-to-death acts as a crude correction for health improvement
- For the OECD exUS we do not have time-to-death, so *D* is just spending in age-sex cells weighted by population share
  - Moreover, we only have spending by age-sex for a single year (2015) for 8 countries, so we assume those spending relatives by age-sex apply to all 19 OECD exUS countries in all years*
- By definition the coefficient of *D*, β_D, is 1.0

*Some comfort is available; for those 8 countries our results are robust whether we use the own-country measure or the 8-country average.

The Relative Price Measure *P*

- Prices are correlated with income and were not included in the specification of Equations 2.1 and 2.2 above because of difficulties in measuring the price variable
- Omitting P will therefore bias β_y; we adjust for the bias
- We want to measure the difference between a medical care price index and an economy-wide price index, but we only have a medical care price index for the US and even that is subject to considerable error for years prior to 1996*

*Berndt, et al. in the Handbook of Health Economics, vol. 1.

The Relative Price Measure *P*, cont.

- By 1996 there were US Producer Price Indices for all medical care sectors, which we use for 1996-2019; for 1970-1995 we impute a medical care price index for the US

Correcting Estimated Income Elasticities from Omitting Relative Price

- Outside the US we lack data to measure *P* directly
- Because price is correlated with income, however, omitting *P* biases the income coefficient, so we use a proxy for *P* based on the Baumol cost disease model*
- Baumol assumes productivity is hard to improve in health care but that health wages must compete with other sectors
  - This implies relative medical price is a function of the differential between economy-wide wage growth and economy-wide productivity growth

*William J. Baumol, *American Economic Review*, June 1967 and *The Cost Disease: Why Computers Get Cheaper and Health Care Doesn't*, Yale University Press, 2012.

Correcting Estimated Income Elasticities from Omitting Relative Price, cont.

- The greater are productivity gains in the progressive sector(s) relative to non-progressive sectors, which is where Baumol places health care, the more the progressive sector price will fall relative to non-progressive sectors
- Thus, we define $\ln(c_h) = \lambda \ln(w/y)$

  where c_h = unit costs of production in health care = (by assumption) output price; w = real compensation per employed person (economy-wide); and y = real GDP per employed person (economy-wide)

The Relative Price Measure for the OECD Excluding the US, cont.

- We have data on ln(w/y) for 14 of our 20 countries and substitute that in our estimated equation for the missing relative price variable for the non-US countries to derive an adjustment for the bias in the estimated income coefficient

The Relative Price Measure for the OECD exUS, cont.

- The results below show the effect on β and β' of adding ln(w/y) to the equation estimated earlier:*

Step 2:	**Estimation of income x relative medical price interaction effect, 14-country sample[3], 1970-2019**				
Estimated coefficient (standard error)[2]	β_y^b	λ	$\beta_y^{b\prime}$	λ'	
(a) Excluding Baumol variable[4] (standard error)[2]	0.7165 (0.0489)		1.3188 (0.0236)		
(b) Including Baumol variable (standard error)[2]	0.6333 (0.0567)	0.6720 (0.0958)	1.2813 (0.0342)	0.3898 (0.1348)	
Adjustment ratio =row (b)/row (a)	0.8838		0.9716		

*The first two columns of numbers show results including year fixed effects and the second two columns show results excluding them. The λ's are the estimated coefficients for ln(w/y).

Quality Change

- Baumol assumes a constant product; his canonical examples are haircuts and live chamber music performances
- A utility-based cost-of-living index should adjust for quality change; two recent papers* suggest doing so would lower US medical price indices in the 1999-2017 period 1.5%
  - The two papers don't test for a break in this period
  - It isn't clear we should use a utility-based price index to explain the medical care share of GDP, but the results in the prior slide in effect adjust for any differential productivity in medical care

*Cutler, et al., *AER*, February 2022, Dunn, et al., *Econometrica*, March 2022.

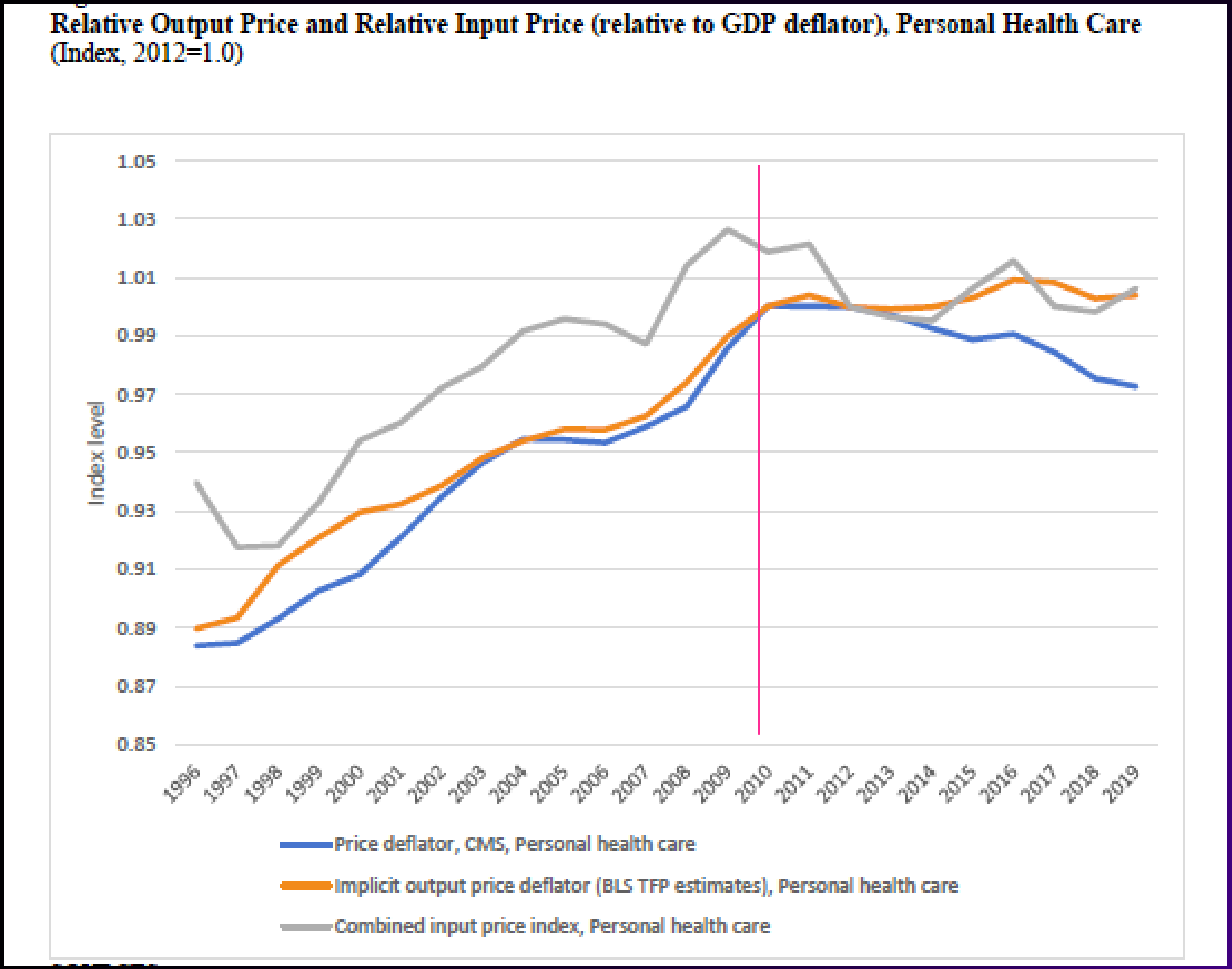
Quality Change, cont.

- If we use Cutler's result that US price indices are overstated by 1.5% annually because they fail to account for quality change, by definition real health care spending increases more rapidly than in the results above because the deflator is less, but our main findings for 1970-2019 are qualitatively unchanged
  - In our decomposition, most of this effect shows up as a larger proportional increase in the role of exogenous technology

Looking Under the Hood of the Negative Effect of Relative Price Inflation

- Partly the negative effect is attributable to the reduction in Medicare reimbursements in the Affordable Care Act
- Exploring further, output price changes equal input price changes minus total factor productivity (TFP) ± margin changes; this is the price dual of the equation for TFP
  - Over a period of several years changes in margins should be small relative to changes in input prices and productivity, so we ignore them

Input and Output Prices Move Together, 1996-2019



An Input Price Index (Gray) and two Output Price Indices all rise before 2010 and flatten after 2010. This suggests the change in the relative price effect after 2009 was driven by changes in input prices and not TFP. The change in input prices after 2010 was comprised of a substantial change in capital prices for hospitals and nursing homes and a smaller change in wages.

Relative Price Effects in the Long Term

- Because the flattening of the rate of increase in output prices *P* after 2010 was (mostly) attributable to changes in input prices and not to changes in TFP, the flattening will probably not be sustained
  - In the long run input price changes in health care should track economy-wide input price changes and thus should not have much effect on share of GDP in health
- By contrast, a change in TFP might have signaled a more permanent reduction in the rate of growth of GDP share