HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

Does One Medicare Fit All? The Economics of Uniform Health Insurance Benefits

Faculty Research Working Paper Series

Mark Shepard
Harvard Kennedy School

Katherine Baicker
University of Chicago

Jonathan Skinner
Dartmouth College

November 2019

RWP19-036

Visit the **HKS Faculty Research Working Paper Series** at:
https://www.hks.harvard.edu/research-insights/publications?f%5B0%5D=publication_types%3A121

The views expressed in the **HKS Faculty Research Working Paper Series** are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. Faculty Research Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

www.hks.harvard.edu

Does One Medicare Fit All?
The Economics of Uniform Health Insurance Benefits

Mark Shepard
Harvard Kennedy School and NBER

Katherine Baicker
University of Chicago and NBER

Jonathan Skinner
Dartmouth College and NBER

November 11, 2019

Abstract

There is increasing interest in expanding Medicare health insurance coverage in the U.S., but it is not clear whether the current program is the right foundation on which to build. Traditional Medicare covers a uniform set of benefits for all income groups and provides more generous access to providers and new treatments than public programs in other developed countries. We develop an economic framework to assess the efficiency and equity tradeoffs involved with reforming this generous, uniform structure. We argue that three major shifts make a uniform design less efficient today than when Medicare began in 1965. First, rising income inequality makes it more difficult to design a single plan that serves the needs of both higher- and lower-income people. Second, the dramatic expansion of expensive medical technology means that a generous program increasingly crowds out other public programs valued by the poor and middle class. Finally, as medical spending rises, the tax-financing of the system creates mounting economic costs and increasingly untenable policy constraints. These forces motivate reforms that shift towards a more basic public benefit that individuals can "top-up" with private spending. If combined with an increase in other progressive transfers, such a reform could improve efficiency and reduce public spending while benefiting low income populations.

∗ We are grateful to Victor Fuchs, Carol Carter, Jeff Clemens, and participants at the NBER Tax Policy and the Economy Conference for helpful comments, and to the National Institute on Aging (T32-AG000186, P01AG19783, P01AG032952) for financial support. Shepard gratefully acknowledges funding from the National Science Foundation Graduate Research Fellowship. Disclosures: Skinner has consulted for Sutter Health, and Baicker is a Director of Eli Lilly and HMS and a member of the CBO and NIHCM Advisory Boards.

1 Introduction

The United States spends 18 percent of GDP on health care, and, at 6% of GDP, federal health spending represents a quarter of the federal budget. With baby-boomers reaching retirement age and the continuing development of expensive new medical treatments, the Congressional Budget Office (CBO) projects a more than doubling of Medicare spending in the next decade, from $711 billion in 2018 to $1.5 trillion in 2029 (CBO, 2019). Recent proposals to extend Medicare to new populations – ranging from lowering the eligibility age to "Medicare for All" – further raise the stakes of using the current Medicare program structure for government-funded health insurance.

Most countries face similar pressures arising from an aging population and medical technology growth, but the traditional Medicare program has specific features that distinguish it from many other insurance plans. It provides a uniform benefit to all enrollees that places few limits on the scope of coverage, even for unproven technologies. By contrast, many countries such as England limit access to new treatments and technologies based explicitly or implicitly on estimates of cost-effectiveness (e.g., Thorlby and Arora, 2019). Unlike most employer-sponsored commercial plans, the traditional fee-for-service Medicare also has no network restrictions on providers, nor (with the wrap-around plans held by most enrollees) significant deductibles or copayments.

In this paper, we focus on the design of the current Medicare program for the elderly to assess its tradeoffs and provide insights about the implications of using it as a foundation for expanding coverage. We first ask about the efficiency and equity tradeoffs involved with its current generous, uniform design. Second, we address the question of how rising income inequality, ongoing medical technology innovation, and the budget pressures imposed by an aging population affect the efficiency of the current benefit structure. Finally, we examine the effects of an alternative, non-uniform benefit structure on economic efficiency and equity.

To study these questions, we build on a rich literature in health economics and social insurance design to develop a simple economic model of Medicare that incorporates income inequality, medical technology growth, and distortionary taxes.[1] The model allows us to assess how the

[1] It is worth noting that our model abstracts from many aspects of the complex US health care system, including consolidated provider and insurance markets, inefficiencies in insurance and payment design, imperfect patient information, and of the underuse and overuse of health care treatments. There are many potentially productive reforms across these dimensions that complement the fundamental Medicare reform described here.

welfare consequences of Medicare's uniform benefit structure have evolved, as well as the welfare effects of potential alternative public insurance designs. We derive predictions using both a stylized graphical framework and simulations from a calibrated version of the model.

The model suggests that while Medicare's uniform benefit has the advantages of simplicity and lower administrative costs, it also comes with a *cost of uniformity.* While high-income households would likely prefer a very generous plan, low-income households would likely prefer lower health care spending and higher take-home pay or more generous non-medical benefits such as food stamps or housing assistance (Baicker, 2001). A uniform program pools everyone into the same plan, creating an inefficiency due to mismatch between the public benefit and privately optimal generosity. This cost of uniformity is closely related to the standard efficiency loss from in-kind transfers relative to cash transfers (Currie and Gahvari, 2008), but it remains in our model that includes an explicit role for in-kind transfers to provide more equitable access to health care.

Our central argument is that three macro trends have increased this cost of uniformity appreciably since Medicare's creation in 1965. First, income inequality has risen substantially (Piketty and Saez, 2014). Rising inequality leads to growing divergence between rich and poor in willingness (and ability) to pay for generous medical care. Second, there have been dramatic innovations in medical technology: there was much less health care available to buy in the 1960s, and even advanced technologies of the day were relatively inexpensive. Third, average marginal tax rates have increased from less than 25% in 1965 to 30% in 2012 (Mertens and Olea, 2018), commensurately increasing the deadweight loss (or economic cost) associated with publicly financed benefits – a trend that will likely continue with the budget pressures from population aging (Baicker, Shepard, and Skinner, 2013).

These changes imply that demand among the rich for generous medical care increasingly diverges from what a uniform public system can afford to fund. While a universal, generous Medicare program may have been efficient in 1965 when options for treatment were both limited and relatively inexpensive, tax rates were lower, and income more evenly distributed, the efficiency cost of maintaining uniform coverage has grown over time. The current benefit design thus may not be a sustainable foundation upon which to expand public health insurance.

We describe an alternative insurance benefit design in which the government provides basic insurance but allows higher-income households to "top up" by purchasing additional coverage for additional services. (Medicare does have in place supplemental "Medigap" plans, but these are

primarily designed to cover copayments and deductibles, rather than cover additional services.) The basic plan is intended to be similar to public insurance provided in many other countries – with low patient copayments and deductibles but with more modest provider payment rates and with coverage of treatments restricted to those with proven effectiveness relative to lower-cost alternatives.

Supplemental "top-up" plans are also common in other countries. Governments often underwrite a basic insurance plan (or mandate the purchase of regulated and subsidized private plans), but then allow households to add on private supplemental insurance.[2] For example, while Swiss citizens are required by law to have basic health insurance, discretionary private insurance accounts for about one-third of total health care spending (Sturny, 2019). In Australia, private insurance offers access to a wider choice of hospitals and faster access to discretionary services (Glover, 2019), while in England 10.5 percent of the population opts for private health insurance coverage (Thorlby and Arora, 2019).

Our calibrated model suggests that switching from a uniform Medicare benefit to a top-up structure could generate substantial cost reductions and efficiency gains in the long term. The distributional implications of such a policy change would depend on the alternative uses to which the resources saved on public insurance would be devoted. Many European countries spend substantially more on other social insurance programs than the U.S., and some of those non-medical programs themselves are likely to yield health benefits (e.g. Bradley and Taylor, 2013; Baicker et al., 2012; Papanicolas et al., 2019). We show in the model that there exists a redistribution of the "Medicare dividend" that would raise wellbeing across all income groups.

While our model considers benefit design solely for people age 65 and over, the implications for Medicare benefit design are clearly amplified under proposals to expand the eligible population. For example, while the cost of "Medicare for All" proposals depends crucially on the details of eligibility, coverage, and provider payment rates, most proposals require additional tax revenues that would substantially raise marginal tax rates.[3] The implication of our simple model

[2] Some countries like Germany allow citizens to "opt out" of the public system and purchase private insurance for their primary coverage. However, under the German system, those with private coverage (and their employers) need no longer contribute to the public plan; for this reason, the German health insurance system more closely resembles a top-up structure

[3] For example, the 2016 Bernie Sanders tax proposal to pay for expanding Medicare included a top federal rate of 52%, with state and local taxes added on. There are of course alternative approaches for generating new tax revenues that come with their own deadweight loss or efficiency costs.)

is that a more basic public benefit – closer to "Medicaid for All" than to "Medicare for All" – with the option for individuals to top up to more generous private coverage, coupled with increased transfers to the poor, could prove to be a higher-value, more sustainable alternative to many proposals that seek to expand the current Medicare program.

2 Health Care Demand and In-Kind Transfers

The public sector plays an outsized role in the financing and provision of health care in nearly all developed countries. Health care is often considered a merit good that ought to be available to all – although often without consensus on the quantity or quality of care to which all should be entitled. But health care is often privately produced, with economic forces driving supply and heterogeneous patient preferences and needs driving demand. These forces have important implications for the optimal design of public insurance programs.

In this section, we present a stylized model of demand for health care and how it varies across income groups based on ability to pay. We then consider how this heterogeneous demand interacts with the design of a public health insurance system. We highlight two key design questions: (1) how generous should the public benefit be, and (2) should it be a fixed *uniform benefit* for all recipients or a *basic benefit* that recipients can *top-up* using their own money?

It is important to define here what we mean by a basic benefit. In the Affordable Care Act health insurance exchanges, "bronze" coverage entails the lowest premium but exposes the enrollee to substantial out-of-pocket payments through deductibles and coinsurance. This is not what we have in mind as our basic benefit, since cost-sharing of this magnitude places considerable financial stress on low-income households (Kullgren et al., 2010). Instead, we define our basic plans as those like in European countries where deductibles and copayments are small, amenities are limited, and provider payments are regulated.

While Medicare hospital-based reimbursement rates are estimated to average just 53% of commercial rates in the US (Maeda and Nelson, 2017; also see Cooper et al., 2018), data from the International Federation of Health Plans (2015) suggests that Medicare reimbursements for

hospital-based and diagnostic services are generally higher than insurance payments in other countries.[4]

Other countries also use cost-effectiveness criteria in restricting expensive treatments. For example, proton beam therapy is valuable for certain childhood cancers and other rare cancers in adults, but there is no evidence that it is more effective than less costly alternatives for the treatment of prostate cancer (Kamran et al., 2019; Schroeck et al., 2017). Medicare covers proton beam therapy for prostate cancer, but England, Canada, and even many U.S. commercial carriers do not.[5]

In this section, we develop a simple model that allows us to consider the equity and efficiency tradeoffs in the design of a public health insurance benefit. In Section 3, we consider how the tradeoffs interact with the three macro forces with which we motivated the paper: rising inequality, expensive new health care technology, and rising tax rates.

2.1 A Simple Model of Health Care Demand with Public Benefits

Consider a stylized model of individuals' consumption (*c*) and medical care (*m*) choices, as in past work such as Hall and Jones (2007). A population of individuals $(i = 1,...,N)$ vary in their private income y_i. Individuals derive flow utility from their consumption $u(c_i)$, and medical spending determines their life expectancy, $\lambda(m_i)$, where we assume diminishing returns to more health care spending.[6] The government provides each person with a cash benefit, *R*, and an in-kind medical benefit, *M*, funded by taxes collected according to a progressive tax schedule $T(y_i)$.

---

[4] We used a report from the International Federation of Health Plans (2015) to make an illustrative comparison of the U.S with four countries (Australia, New Zealand, Spain, and Switzerland). We examined 9 treatments or diagnostic procedures: appendectomy, bypass surgery, cardiac catheterization, cataract surgery, colonoscopy, CT scan (abdomen), hip replacement, and knee replacement. The prices in the report represent a combination of public and private insurance plans. We assume Medicare reimbursement rates of 53 percent of commercial US prices (following Maeda and Nelson, 2017 and Cooper et al., 2018). This calculation suggests that average Medicare prices were at least 10 percent higher than those in any of the other countries. That said, there was considerable variation across countries and procedures in relative prices, with some lower-cost procedures having lower prices in Medicare than in other countries. Similarly, a recent study estimated that Medicare pays 80 percent more for prescription drugs than insurance plans in other high-income countries (ASPE, 2018).

[5] For example, the Prostate Cancer Canada (2019) website states that "Because proton therapy is very expensive and has no significant advantages in cure rate for prostate cancer compared with other treatments, it is not covered by Canadian public health insurance – you would need to pay for the procedure yourself." Also see National Health Service (2018). There are exceptions; the Netherlands, which has invested heavily in proton beam facilities, does cover their use for prostate cancer.

[6] It is straightforward to interpret $\lambda(m_i)$ as quality-adjusted life-years; that is, health care inputs m are likely to affect the quality as well as quantity of life.

Individuals choose whether to buy additional top-up care m_i subject to constraints imposed by the system. We write the individual choice problem as:

$$\begin{aligned} &\max_{c_i, m_i} \lambda\left(M + m_i\right) \cdot u\left(c_i\right) \\ &\text{s.t.} \quad c_i + m_i = y_i - T\left(y_i\right) + R \end{aligned} \tag{1}$$

The constraint on m_i depends on the design of public insurance. We consider two possibilities:

1. **Uniform benefit:** $m_i = 0$ *(individuals cannot top-up the public benefit)*
2. **Top-up benefit:** $m_i \geq 0$ *(individuals can top-up above M with private funds)*

Note that this setup also allows for a purely private health system, which is equivalent to a top-up benefit with $M = 0$.

Now consider demand for health care under a completely private system ($M = 0$). The model predicts, not surprisingly, that privately chosen medical spending rises with income. That is, health care spending equates the value or marginal willingness to pay (WTP) to extend life by an additional year, $WTP_{LY}\left(c\right)$, with the marginal cost of doing so, $MC_{LY}\left(m\right)$. This can be seen from the first-order condition for *m*, which can be expressed as:

$$WTP_{LY}\left(c_i\right) \equiv \frac{u\left(c_i\right)}{u'\left(c_i\right)} = \frac{\lambda\left(m_i\right)}{\lambda'\left(m_i\right)} \equiv MC_{LY}\left(m_i\right) \tag{2}$$

A key insight of Hall and Jones (2007) is that willingness to pay to extend life rises steeply with income; for the wealthiest in society, the marginal value of another Lamborghini is low, but an additional year of good health in which to enjoy it is nearly priceless.[7] As a result, privately chosen medical spending also increases rapidly with income.[8] Relative to the poor, the rich proceed further up the marginal cost curve, spending more on less valuable services until $MC_{LY}\left(m\right)$ equals their much higher value of a life-year.

In Panel A of Figure 1, we graphically depict an example of this private health care-income relationship in a simple case with two income types, rich (*H*) and poor (*L*). The x-axis is medical spending (*m*), and the green curve is the marginal cost of a life-year (MC_{LY}), as defined above.

[7] This is evident from the expression for WTP of a life year: higher income people have higher consumption, implying a larger $u\left(c_i\right)$ and smaller $u'\left(c_i\right)$.

[8] This does not need to require that health care is a "luxury" good (income elasticity > 1) but simply that it is a normal good (income elasticity > 0). See Acemoglu et al. (2013) who argue that health care is normal but not a luxury.

For a given level of medical technology, marginal cost first rises gradually but then steeply as the limits of medical technology are reached, or as more spending takes the form of amenities. Unlike a physical production technology, at some point one can no longer "produce" additional units of health simply by spending more; when all possible treatments are exhausted, the marginal cost curve is vertical. [9]

The WTP for an additional life-year is depicted by the downward sloping blue curve ("WTP_{LY} *(Rich)*") for the rich type and the orange curve ("WTP_{LY} *(Poor)*") for the low-income type. (We defer discussion of the dashed orange curve to the next subsection.) Because willingness to pay is much higher for the rich, the level of medical spending is greater for the rich (m_H) than the poor type (m_L).

For simplicity, we include just a single input measure, *m*, but in practice demand for health care across income groups may also reflect differences in types of treatment, convenience, and amenities not immediately reflected in survival. For example, Cheng et al. (2018) found that lottery winners in the United Kingdom (i.e., those experiencing an unexpected increase in income) did not consume more health care in terms of hospital days or physician visits but instead shifted to private rather than public care. This finding is in line with how we are thinking of "generosity" in the model – as improving quality (through short wait-lists, more comfortable hospital beds, and greater provider choice) rather than quantity.

An entirely free market in health care would be viewed as inequitable if the level of care the poor can afford is below what society has deemed adequate. Cash transfers can narrow the gap, but most countries use in-kind health insurance benefits as a more direct way of doing so – most likely because of a deep-seated concern about the health of their fellow citizens, or what we call

[9] An anecdote illustrates this point. An ICU physician, Goetz (2004) stated in a letter to *Health Affairs*: "Here is an example I have used when teaching medical students and residents: You are taking care of a patient in the ICU. You have done every test and procedure you know to do and have done everything that all the consultants have recommended. I now tell you that you must spend another $5,000 (originally I used $1,000) to improve the patient's quality of care. What would you do with the money? By this point the student or resident is in a bit of a quandary because they are not quite sure how to use the additional money. If there were a continuing positive linear relationship, it should be reasonably easy to suggest more things that result in improved patient care. Generally, the suggestions are more, or repeated, tests and procedures. I respond to the common answers with a statement that if you do more tests or procedures, you could in fact make the patient worse. How? If you do more tests, all tests have false positives and negatives. How will you use results that contradict earlier tests? With again more tests, and the subsequent potential for much more confusion. If you repeat or do another procedure, how do you interpret the results? Also, procedures generally have potential side effects or complications, so again you have a very high risk of NOT improving quality or outcome with more money."

an *egalitarian social preference* for health. We turn next to study how such egalitarian preferences affect the optimal benefit design for health insurance.

2.2 Egalitarian Social Welfare and the Government Problem

Absent political constraints, the government has considerable latitude both in designing how transfer programs are funded and in choosing the composition of benefits between in-kind medical care and cash (or near-cash) benefits. The government can levy taxes to fund cash and in-kind transfers. But why provide in-kind health insurance benefits at all? The argument in favor of cash benefits is familiar from Economics 101: for the same cost to taxpayers as an in-kind benefit, cash benefits raise recipients' utility more because they do not constrain their choices. But there are multiple arguments in favor of in-kind benefits, such as better differentiation between targeted and non-targeted beneficiaries and taxpayer preferences about recipients' use of resources (Currie and Gahvari, 2008).

We focus on a rationale based on *egalitarian social preferences* for an equitable distribution of health care.[10] A simple way to capture the idea of egalitarian social preference is to add to private utility an additional term, $e(m_i)$, capturing the value to society of an individual *i*'s access to medical care – or the extent to which taxpayers care about the health care of others. We assume that $e'(m) > 0$ for low levels of spending but that it declines and reaches zero at some "adequate" level of health spending – such that for the high-income person in our graphical analysis, it has dropped to zero. Returning to Figure 1 Panel A, this egalitarian preference shifts out the social value of health care for the poor type from WTP_{LY}^{Poor} up to the dashed orange line. Society would like the *L* type to obtain medical care $m_L' > m_L$. The divergence between private and social optima motivates the in-kind medical benefit.

The government funds these benefits with taxes. To allow for progressivity in a simple framework, we model taxes as a scaled version of a baseline progressive schedule $T_0(y_i)$, so that $T(y_i) = \tau \cdot T_0(y_i)$ where τ is the scale factor. Rather than model the full labor supply problem,

[10] Other rationales suggested in the literature include direct externalities from the in-kind good, indirect benefits through the tax/transfer system (e.g., subsidizing goods that complement work, or providing goods differentially attractive to low-ability types), and insurance market failures such as adverse selection; see Currie and Gahvari (2008). We focus on equity, which seems the most straightforward way to justify a uniform program like Medicare.

we model deadweight loss in a simple way. We assume that levying taxes that reduce individuals' consumption by \$1 only raises tax revenue of $\$\left(1-\chi(\tau)\right)$, where $\chi(\tau)$ captures revenue leakage due to the excess burden of taxes. As taxes rise, $\chi(\tau)$ will increase due to the rising marginal excess burden of higher tax rates.

Putting these pieces together, the government chooses the tax scalar (τ), cash transfer (R), medical benefit (M), and design of medical benefits to maximize the social welfare function:

$$SW = \sum_i \left[\lambda(m_i)u(c_i) + e(m_i)\right] \tag{3}$$

subject to public budget constraint:

$$\sum_i \left(R + M\cdot(1+\kappa)\right) = \sum_i \tau \cdot T_0(y_i)\cdot\left(1-\chi(\tau)\right)$$

and where $\{m_i, c_i\}$ are set by individual choices from the problem in (1) and κ captures the excess administrative costs of providing in-kind medical benefits.

Benefit Design: Uniform vs. Top-Up

The government can provide medical benefits with either a uniform or top-up design. It is natural to ask why one would ever adopt a uniform benefit design instead of a top-up design. A uniform benefit imposes an additional constraint on recipients – a ceiling on medical spending at M – not present with top-up benefits. However, in practice, a top-up system creates additional complexity, including greater administrative costs and adverse selection problems associated with letting consumers choose among generosity levels. We model this by assuming that $\kappa^{TopUp} > \kappa^{Uniform}$ in the problem above. The lower κ provides a rationale for the uniform system, all else equal.

In the case of a uniform benefit, the government must decide on a level of care that balances the different demands for the lower income households and for the high income households. One seemingly egalitarian solution would be to make the uniform level of coverage equal to m_H, so that few, if any, high-income households would want to opt-out into a more generous private plan. This appears to be what we observe with the U.S. Medicare program. Yet the downside of providing the generous plan is the opportunity cost of doing so, including less ability to fund cash transfers or other social programs that might provide greater value to lower-income households.

The economic model presented above suggests an alternate way of setting an optimal uniform public medical benefit: finding a middle ground between the ideal points for the rich and poor. Panel B of Figure 1 shows this middle-ground benefit, M^*_{Unif}, which is less generous than what the rich would like (m_H) but more generous than the social optimum for the poor (m'_L).[11] This result is analogous to the Samuelson condition for an optimal public good: it is set where the average WTP of beneficiaries is equal to the average marginal cost of providing the service.

Even when chosen optimally, there is an implied efficiency loss as shown in the red highlighted areas in Panel B of Figure 1. High-income households would prefer more than what was provided, while low-income households demand less, meaning that they might prefer the cash to high-amenity health care or access to unproven treatments. We call this loss the *"cost of uniformity"* involved with a uniform benefit system.

Figure 2 plots the optimal determination of top-up benefits (superimposed on top of the optimal uniform benefit from Figure 1B). The top-up design still provides a public benefit to ensure a floor on medical spending for the poor, but it does not impose a medical ceiling on the rich. This allows the government to set the optimal top-up benefit (M^*_{Top})based on the social optimal level for the poor – where the marginal cost curve intersects the social WTP for a life-year. The rich top-up by purchasing health care privately up to their desired value. In this simple two-type case, the top-up system eliminates the "cost of uniformity" by letting each group get their optimal level of care.[12] However, it also creates additional administrative/complexity costs (κ^{Top}), shown by the upward shift of the green marginal cost curve and the highlighted losses in the green area. We turn next to a more detailed discussion of the tradeoffs involved between the two systems.

2.3 Tradeoffs between Uniform vs. Top-Up Medical Systems

Notice several implications of optimal top-up benefits in comparison to the uniform design, as depicted in Figure 2:

- ***Public* health insurance benefits are lower under top-up benefits** – (i.e., $M^*_{Top} < M^*_{Unif}$)

  This result follows from the logic of the public goods problem. The uniform benefit

[11] Technically, the optimum occurs where the curve is vertically equidistant from the blue curve and dashed orange "social value for poor" curve.

[12] In a more realistic case with many types, the cost of uniformity would not be completely eliminated, since any floor created by a public benefit would still constrain the choices of some people with low willingness to pay.

determines medical spending for everyone, while the top-up benefit determines medical spending only for the (poorer) constrained group who choose not to top-up. Eliminating uniformity frees up the government to set up a low-cost, basic public benefit based on demand by poorer households, rather than having to balance the desires of the rich and poor in a single system, or (as Medicare appears to be designed) a uniform program targeted to the needs of higher income households.

- **However, *total* medical spending (public and private) may be higher or lower in the top-up design.** It will be lower for poor households (who receive a smaller benefit) but larger for the rich (who top-up), so the overall change depends on the shape of the marginal cost curve and the size of each group in the population.

- While not evident from the graphs, under the parameters we assume in the model, **optimal *cash transfers* can be higher under a top-up medical system.** Intuitively, cash and health insurance benefits are substitutable forms of redistribution. As noted by Baicker (2000) and Bradley and Taylor (2013), when the government spends less on health care, there is more to redistribute as cash (or use for other purposes). This fact is important for interpreting the equity implications of the two designs. While the poor get less generous health care under top-up benefits, they may also get more cash income.[13]

- **The top-up system (by design) allows for greater medical inequality.** This health care disparity for rich vs. poor is often seen as inequitable – even if it originates from private choices and allows the government to fund more cash transfers.

Whether a uniform vs. top-up design is preferred depends on the relative losses from uniformity (red areas) versus losses from administrative complexity (green area). We discuss next the reasons that this cost of uniformity has likely risen over time.

[13] This result depends on the preferences of taxpayers; if they only care about health care for the poor and nothing else, they may not choose to distribute more in cash.

3 Rising Cost of Uniformity: Medical Technology, Inequality, and Taxes

As noted above, uniform benefits provide equal access to health care, but this uniformity comes at a cost when income groups differ in their demand for health care. In this section, we ask how this "cost of uniformity" has changed over time. The medical and economic world was quite different in 1965 when Medicare was created. How should we think about the impact of forces like improved medical technology, rising income inequality, and changing tax rates? Using the graphical framework developed above, we argue that these forces tend to imply *rising costs of uniformity*.

3.1 Baseline: Medicare in 1965

The four panels of Figure 3 walk through the basic logic. Panel A starts by applying the model to the environment at Medicare's founding in 1965. There was much less income inequality than there is today, reflected in the narrower gap between rich and poor's WTP for medical care. Medical technology was much less advanced (and expensive) than today. In 1965, few options for treatment were available for common health care conditions. Rather than a production function with gradually diminishing returns, there was relatively little to do beyond relatively few low-cost interventions for common diseases such as cancer and cardiovascular disease.[14] Graphically, this is reflected in the steep asymptote in the green marginal cost curve per life-year at a relatively low level of medical spending.

The key implication of this 1965 environment is that optimal medical spending for rich and poor were not too divergent. Just as today, the rich could certainly afford more and better health care than the poor. But after accounting for the egalitarian social value of providing basic health care to the poor, the optimal spending for rich and poor (labeled as m_H and m_L') are quite close. As a result, optimal uniform benefits (M^*_{Unif}) are also not too different from either group's private optimum and the loss from the uniform constraint (red area) is small. This logic suggests that when Medicare was established, a single uniform program for rich and poor seniors made good sense. The cost of uniformity was low, so even a small complexity cost from a more flexible top-up

[14] Bypass surgery had been developed by 1965, but this was a high-risk and quite rare intervention that was not appropriate for the vast majority of people with cardiovascular disease; statins, stents, and trans-catheter aortic valve replacements (TAVR) were still far in the future.

system would be enough to tip the scales towards a uniform program. Moreover, this program would optimally be quite generous, since health care was relatively cheap and the government budget (and associated taxes) relatively small.

3.2 Improving Medical Technology

Now consider the impact of improving medical technology from new and more expensive treatments (Cutler et al., 2006; Chandra and Skinner, 2012; and Howard et al., 2015). Treatments for nearly all conditions – but especially heart disease and cancer – have vastly advanced since 1965, but these new treatments are also very expensive. The heart attack that would have resulted in death in 1965 can now be treated and a life extended – but at a cost per hospital admission of $20,000 or more.

Panel B in Figure 3 shows how improving medical technology is reflected in our graphical model. The result is a large outward shift and flattening of the green marginal cost curve. This shift reflects the idea that at a given level of *m*, the marginal returns to medical spending are much higher than 1965. The marginal cost curve (or dollars per life-year) is the reciprocal of the marginal returns (e.g., life-years per dollar), so the marginal cost curve is lower and flatter. Of course, the marginal cost curve eventually steepens at a much higher level of *m* – and these are the marginal technologies that are both high-cost and low-value.

The outward shift and flattening of the MC curve have an important implication: Privately demanded medical spending for rich and poor are now much further apart. This follows from the simple logic of demand that higher-income households are more willing (and able) to pay the huge bills associated with modern medicine. Indeed, Medicare's annual costs of $10,739 per elderly enrollee is 63% of the average Social Security benefit ($16,956 in 2017), which is the only source of income for many lower-income seniors.[15] The poor simply cannot afford to pay the full cost of the Medicare premium without a substantial subsidy.

The egalitarian social value narrows but does not close this growing gap between rich and poor. As a result, the optimal uniform benefit (M^*_{Unif}) is much further apart from the group-specific

[15]For Medicare spending in 2017, see https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nhe-fact-sheet.html For the average Social Security benefit in the same year, see https://www.fool.com/retirement/2017/08/30/how-big-is-the-average-persons-social-security-che.aspx

optima for rich and poor (m_H and m_L') than in the baseline in Panel A. This divergence implies a larger cost of uniformity, reflected in the bigger red shaded area in Panel B.

3.3 Rising Income Inequality

One of the major economic trends of the past 50 years is rising income inequality (Piketty and Saez, 2014). While low and middle-income earners have seen modest economic growth since about 1975, the rich have experienced rapid gains. Rising inequality should also be reflected in rising inequality in demand for medical care.

We depict rising inequality in Panel C of Figure 3 via a large upward shift in the willingness to pay (WTP) for a life-year for the rich. (Note that we retain the marginal cost curve for 2015 from Panel B.) The rich therefore proceed much further up the marginal cost curve up to the point that it starts becoming quite steep – reflecting the marginal low-value care. For example, private hospital rooms or proton beam therapy for prostate cancer may be more convenient and worth the extra expense for high-income households, but given the lack of demonstrated health benefits, might not be worth it for low-income households.[16] Indeed, the example of proton beam therapy shows how rising inequality works together with expensive new medical technologies to produce the patterns we describe, since rising income inequality would have little impact on medical spending gaps without the existence of expensive treatments with limited health benefits.

Therefore, rising income inequality – in conjunction with improving technology – results in a growing divergence in optimal medical spending between rich and poor and a growing divergence between the optimal top-up and uniform benefits. The red shaded "cost of uniformity" grows larger versus Panel B. Further, the rich are increasingly dissatisfied with the care provided by the uniform medical benefit, which falls short of the generosity they would privately choose.

The poor, while not dissatisfied with the generous medical care itself, could increasingly prefer less generous medical care (perhaps closer to Medicaid in coverage) coupled with assistance in paying for other aspects of health care, including Part B premiums (at least $1,626 annually in traditional Medicare for those not covered by Medicaid), and assistance with other non-medical needs such as housing, transportation, or their grandchildren's educational debt. This is the

[16] Other examples of tradeoffs between costs and convenience/quality come from comparing Medicaid to Medicare. Medicaid is lower cost than Medicare but provides less choice of physicians (about 30% of doctors do not accept Medicaid). Medicaid is more aggressive about managing care, denying claims, and requiring prior authorization for expensive treatments (Gottlieb, Shapiro, and Dunn, 2018).

paradox of the egalitarian motive to provide equitable access to health care; while leveling the health care playing field, it comes at the opportunity cost of forgoing other public assistance that the poor and middle class might prefer. This opportunity cost becomes sharper over time as medical costs rise and inequality grows, making a basic top-up program increasingly attractive even to the non-rich.

3.4 Rising Deadweight Loss of Taxes

Although top income tax rates have fallen since the 1960s, average overall marginal tax rates are higher today than in 1965 (Mertens and Olea, 2018). Moreover, the large federal debt and impending cost of Social Security and Medicare for an aging population suggest that tax rates will likely rise further, which suggests rising deadweight loss of taxation. What effect does this have in our framework? The idea, which we show more formally in our model in Section 4, is that when the marginal tax rate is 50 percent (for example), the efficiency cost (whether through reduced labor supply, capital accumulation, or tax avoidance) of raising an extra dollar of revenue is higher. Thus, publicly provided health insurance must be that much more valuable to society. We depict this in Panel D of Figure 3 via a leftward shift (i.e., reduction) in the optimal uniform benefits.

Although a higher deadweight loss of taxation would also reduce optimal benefits under a top-up system, the welfare impact is larger under a uniform system. With a top-up system, the rich can purchase additional care above the public benefit and obtain privately optimal generosity (m_H) with their own money. But with the uniform system, the rich must consume only the public benefit, moving further away from what the rich would privately choose. This increases the total size of the red shaded area, which was previously set to optimally balance needs of rich and poor but is now more tilted towards the poor. Intuitively, higher deadweight losses from taxation requires cutting the public benefit, which in turn makes the restriction against topping-up in the uniform system more costly.

We have presented a simple graphical model that describes a variety of factors that may affect the desirability of a uniform health insurance plan like Medicare, but one might reasonably ask whether these factors are important in more than a theoretical sense. For this reason, we turn next to a model of health insurance demand across the income distribution (rather than just for rich and poor), using plausible parameters based on the empirical literature and calibrated to fit the observed patterns of health care utilization during the past 50 years.

4 Simulation Model of Uniform vs. Top-Up over Time

We specify a simple simulation model with parameters drawn from the literature and calibrated to economic data to simulate how much secular trends in income inequality, medical care technology, and the marginal cost of taxation affect the relative advantages of a uniform insurance benefit compared to a top-up system.

4.1 Model Setup and Calibration

To judge the magnitude of these effects in practice, we calibrate a model matched to U.S. data with income heterogeneity, income redistribution, technology growth, and an egalitarian preference that values longevity for those with low income. We calibrate the model's parameters to capture the growth in income, income inequality, and medical spending over the 1968-2045 period. (We begin in 1968 to allow the phasing in of Medicare benefits and enrollment.) We then solve the model under different policies at roughly 10-year intervals (1968, 1975, 1985, …, 2045) for a population of 100 individuals who each represent one percentile of the income distribution. We outline the model setup and its calibration here; model details are discussed Appendix A.

The simulation model is based closely on the public policy problem in equation (3) and the individual decision problem in equation (1). To implement this empirically, we need to specify an income distribution and a functional form for consumption utility $u(c)$, the health production function $\lambda(m)$, and the progressive tax function $T(y)$. We discuss each of these pieces in turn.

Income, Taxes, and Deadweight Loss

For the income distribution, we draw on historical data from the Census/ACS for 1970-2015 (with all values inflation-adjusted to 2015 dollars) to estimate the income distribution in each simulation year. This allows us to capture both income growth and widening inequality. We project forward future income distributions from the 2015 distribution using an assumed 1.5% real annual growth rate.

For taxes, we estimate a smooth increasing baseline tax function $T_0(y)$ for the average tax rate matched to the actual schedule for U.S. income, payroll, and state taxes (see Appendix A for a plot of this function). As discussed in Section 2, the government then chooses a tax multiplier (τ) to balance its budget, with the final tax function equaling $T(y_i) = \tau \cdot T_0(y_i)$. As also discussed

in Section 2, for the excess burden of taxes, we do not model the full labor supply problem but instead specify that levying taxes that reduce individuals' consumption by \$1 only raises tax revenue of $\$(1-\chi(\tau))$, where $\chi(\tau)$ captures revenue leakage due to the excess burden of taxes. As taxes rise, $\chi(\tau)$ will increase due to the rising marginal excess burden of higher tax rates.[17]

Utility and Health Production Functions

The next pieces of our model are the utility and health production functions. For utility, we specify a log-utility form with a constant added to match a value of life-year of \$100,000. In addition to this private value of life, there is an egalitarian social externality of health care ($e(m)$ in the model). We specify the egalitarian externality as an additional social value of living an extra year, valued at the flow utility of consuming the year's median income, $u(\bar{y})$. By using a fixed $u(\bar{y})$ for all individuals, it effectively places more weight on improving the health of the poor relative to their private value of health.

For the health production function, we specify a simple function (see Appendix A) that captures the probability of survival as a function of medical spending. A key feature of the function we adopt is that the ability to purchase better health with more spending is *bounded*. With an unbounded function (e.g., a constant elasticity function), any extension to longevity can be purchased with sufficient spending, which often leads to health care being an extreme luxury good, with higher-income people always spending a larger share of their income on health care. A bounded health production function capturing more closely both the medical reality and empirical evidence on income elasticities (Acemouglu et al., 2013).

We then calibrate the model's parameters to match two moments in each simulation year: (1) the historical or projected Medicare spending per beneficiary, drawing on projections from the Medicare trustees, and (2) historical or projected average life expectancy at age 65 from the Social Security Administration (Bell and Miller, 2005). By flexibly calibrating $\lambda(m)$ in each simulation year, we capture the way that technology improves to both extend life and to induce people to

[17] We specify $\chi(\tau)$ to approximate the average revenue leakage from a tax increase from zero up to a given marginal tax rate. See Appendix A for the derivation.

spend more on medical care. Calibrated medical spending rises from 4% of average income in 1968 to 15% in 2015 and to 24% in 2045.

Government Policies

We model government policies as follows. For simplicity, we model a fixed exogenous cash transfer of R = 5% of the year's average income. We also assume "extra" government spending (on non-transfer programs like defense) of E = 10% of average income. With these two policies set, the government makes a single policy choice of the generosity of health insurance benefits (M). Taxes (τ) are set at the level needed to balance the government budget. We consider three policies for health insurance benefits:

1. **Generous uniform:** A uniform benefit (no top-up allowed) with the benefit level exogenously set the level of medical care a rich individual (95th income percentile) would privately choose to purchase; we believe this is a reasonable approximation to Medicare's generous plan.
2. **Optimal uniform:** A uniform benefit with the level of M set to maximize the egalitarian social welfare function. This would correspond to a Medicare-style uniform plan but with likely more restrictive benefits, and with topping-up prohibited.
3. **Optimal top-up:** A basic benefit M that individuals can privately top-up (with $m_i \geq 0$) if they choose. The level of M is set to maximize egalitarian social welfare, recognizing that some individuals will choose to top-up.

We think of the generous uniform policy as roughly analogous to Medicare in that it provides extensive coverage and provider choice so that even higher-income households view the plan as roughly what they would demand even in the absence of Medicare.[18] Because it is an popular entitlement and is not subject to a budget constraint, its generosity evolves mechanically, so that it grows even in the face of budget pressures and changing economic fundamentals. Optimal uniform and top-up benefits, by contrast, reflect the model's optimal tradeoff between equity and efficiency given the respective benefit structures and changing economic circumstances. For top-

[18] Even when wealthy elderly people sign up for "concierge" medical services, Medicare is still commonly relied upon for inpatient services.

up benefits, we assume an additional administrative cost of $\kappa = 2\%$ (with $\kappa = 0$ for the uniform systems), which ensures that the top-up policy need not always be socially preferable.

4.2 Simulation Results

Medical Spending Inequality

A key premise of our analysis is that inequality in *private demand* for medical care across the income distribution has widened over time. Although it is difficult to observe real-world measures of the income elasticity of private medical demand – not influenced by public programs or employment-based health insurance[19] – we can test this idea in our simulation model.

Figure 4 plots privately chosen medical spending for various income percentiles (25th%, 50th%, 75th% and 95th%) in a simulation with cash redistribution (R = 5% of average income) but no public health insurance benefits (M = 0). All values are reported as a share of the average income in each year; this adjusts for general economic growth but makes levels comparable across groups for a given year.

Medical spending rises sharply for all groups over time, reflecting the improved medical technology in our calibrated health production function. But in addition to general growth, *inequality* in health spending across incomes widens. While the rich always buy more health care than the poor, the gap grows as health care becomes more expensive relative to income. In 1968, the 95th income percentile spends 4.1% of mean income on health care (about $2,500) versus 1.4% of mean income (about $880) for the 25th percentile – a gap of 2.7% of mean income. By 2045, these numbers grow to 23.9% of mean income ($31,000) for the 95th percentile versus 9.6% of income ($12,500) for the 25th percentile – a gap that has widened five-fold to 14.2% of mean income. Most of this widening gap reflects the simple fact that medical spending grows as a share of income. As a result, similar *proportional* gaps – the 95th percentile spends about 2-3 times the 25th percentile in all years – implies much larger *absolute* gaps in desired spending.

[19] Studies on this topic have used income shocks such as oil price shocks (Acemoglu et al. 2013) or lottery winnings (Cheng et al. 2018) to estimate positive income elasticities of health spending. Another source of information is individual market health insurance choices in exchanges. Demand estimates from these settings are consistent with higher-income people being less price-sensitive – and therefore more sensitive to quality – in plan choices (Shepard 2016, Jaffe and Shepard 2019, Tebaldi 2017). However, there are few measures on whether or how this income elasticity has changed over time.

Public Health Insurance Benefits

Under a uniform insurance system, the government must choose a single level of health insurance benefits for everyone. But because of rising inequality in desired spending, any single choice will increasingly diverge from the private optimum for many income groups. This presents a dilemma for policymakers. If they choose benefits to satisfy the rich (as in the "generous uniform" policy), this choice will be increasingly inefficient for the middle class and poor who would prefer other forms of support. But if they choose benefits to satisfy the middle class, the rich will become increasingly dissatisfied.

Figure 5 shows how our simulated policymaker balances this tradeoff under the three policies we consider: generous uniform, optimal uniform, and top-up benefits. It plots the chosen level of the public medical benefit (M) over time, both in levels as a share of average income (Panel A) and in terms of its percentile in the private medical distribution from Figure 4 (Panel B). Recall that the generous uniform policy is mechanically set at the 95th percentile of private medical spending, so it grows accordingly. The optimal uniform and top-up benefits also grow but at a slower rate so that they diverge over time from the generous uniform policy.

In 1968, all three policies are similarly generous: 4.2% of mean income for generous uniform, 3.8% for optimal uniform, and 3.6% for top-up – or a gap of 0.6% of income from most to least generous policies. By 2015, the gap has widened, with benefits of 14.7%, 12.4%, and 11.7% of mean income for the three policies – or a gap of 3% of income. By 2045, it has still further widened to 24.0%, 18.5%, and 16.7% of the mean income – a gap of over 7% of mean income. These numbers also indicate that the gap between optimal uniform and top-up benefits grows over time (from 0.2% of income to 1.8% of income), another prediction of our graphical analysis.

Panel B shows this divergence another way. While the generous uniform policy grows with the preferred spending of the rich, both the optimal uniform and top-up policies cannot keep up with this growth. They fall relative to the distribution of privately optimal health spending, from the 90th and 85th percentiles in 1968 down to the 69th and 59th percentiles in 2045. This distributional fall also suggests the growing rationale for a top-up system. In 1968, just the richest 10% of people want health care more generous than the optimal uniform benefit. By 2015 this share grows to 22% and by 2045 to 31%. This suggests that the political economy of a uniform system without a top-up would be increasingly difficult to sustain (Baicker et al. 2013).

Underlying these simulation findings is the changing tradeoff between efficiency and (egalitarian) equity being made by the model. In the years around 1965 when health care was cheap, the cost of providing generous uniform insurance for all was low. But over time, the efficiency cost rises for two reasons. The first is the rising inequality in privately desired health spending highlighted above. A single uniform policy is increasingly divergent from what individuals would have chosen on their own. The second reason is that the rising cost of public health insurance pushes up tax rates, and therefore the marginal deadweight loss of taxes. Generous medical care becomes increasingly unaffordable even for the government. Figure 6 plots the top marginal tax rate under the different policies. This top rate rises from about 30% in 1968 to 66% in 2045 under generous uniform insurance. Shifting to the less generous optimal uniform or top-up benefits allows for a more moderate top marginal rate of about 50% in 2045.

Distributional Incidence and a Progressive Top-Up Reform

How do the uniform and top-up policies affect overall welfare and the welfare of different income groups? In the simulations above with fixed cash benefits, less generous uniform health insurance benefits imply lower taxes, which disproportionately benefits the rich. Moreover, shifting to a top-up system further benefits the rich because it removes a constraint on their desire to buy health care beyond the public benefit. It also allows the government to reduce the generosity of the public benefit, which is reflected in the lower M under top-up in all years in Figure 5A. For these reasons, shifting from generous uniform to either optimal uniform or top-up is at baseline a regressive change.

However, this distributional effect comes from our (simplifying) assumption that the savings from lower public medical spending – or the "Medicare dividend" – is entirely devoted to lowering taxes. In practice, governments have a choice of how to spend this Medicare dividend. Governments could instead use the savings to increase cash transfers, which may allow for a more progressive incidence. We consider a policy exercise in which, starting from the top-up policy modeled above, we increase cash transfers (R) (and the taxes needed to finance them) until *all* individuals in the bottom 30% of the income distribution have higher egalitarian welfare (utility plus the egalitarian externality) than their welfare under the generous uniform policy. If this change can be implemented while still reducing government spending, it suggests that the efficiency rationales for shifting to a top-up system can be realized while simultaneously benefitting the poor.

Figure 7 shows the welfare change across the income distribution due to the shift from generous uniform to the baseline and progressive top-up policies in several simulation years (1975, 2015, and 2035). In each panel, the x-axis is the income percentile, and the y-axis is the (equivalent variation) welfare change for that income level expressed as a percent of income.[20] In all years, the baseline top-up policy (shown in black) reduces welfare for the poorest individuals and increases welfare for the rich. The welfare changes from the baseline top-up policy are small in 1975 but grow sharply by 2015 and 2035. This is consistent with the point discussed above that uniform and top-up policies are in practice quite similar in early years but diverge over time.

The blue curves in the panels of Figure 7 show the welfare change from the progressive top-up policy. By construction, the poorest 30% of incomes all have a positive welfare change. In practice, however, welfare increases for *all income groups* in 2015 and 2035, and in 1975 increases for everyone except some middle-income groups (whose welfare is essentially unchanged, declining by no more than 0.1%). This improvement is paid for by a small reduction in the welfare gains for the richest 30%, who pay higher taxes to fund the increased cash transfer. Despite this, the rich are still the primary beneficiaries of the policy, with welfare gains of 1-5% of income in 2015 and 3-9% of income in 2035. This suggests that governments might consider even larger increases in cash redistribution to share the welfare gains more equitably.

In addition to this, total public spending (shown in the text at the bottom right of each graph) decreases relative to the generous uniform policy. For instance, in 2015 public spending declines by 2.6% of mean income under the progressive top-up policy. This decline is only slightly less than the 2.8% of mean income decrease under baseline top-up, with the 0.2% difference representing the higher cash transfer. Thus, a relatively small amount of cash redistribution is enough to ensure that the poor are not hurt by the shift to top-up in our simulations.

These welfare and spending results illustrate the "win-win" nature of the progressive top-up reform. Underlying these results is the fact that the generous uniform policy – which provides medical care designed for the rich to all income groups – is increasingly inefficient over time. The simple switch to a more basic medical benefit that allows for top-up generates a welfare gain that (if shared equitably) can improve average welfare across the board.

[20] Equivalent variation measures the change in income at the initial policy (generous uniform) required to match the individual's welfare level at the new policy (top-up). We use egalitarian welfare as the welfare metric in this calculation to avoid having our results be driven by the divergence between individual utility and social welfare that is the original motivation for in-kind medical benefits.

5 Discussion and Conclusion

Means-tested in-kind transfers of housing, food, and health care are the predominant form of income redistribution to low-income households (Currie and Gavhari, 2008). Medicare is a prominent example of a uniform in-kind benefit provided to *both* high- and low-income populations. In this paper, we develop a model that allows us to gauge the tradeoffs involved in this uniform benefit design.

Using a stylized model, we show how tax distortions, income inequality, egalitarian preferences, and technology growth affect the efficient structure of the program. Our results suggest that in 1965 when Medicare was first created, its uniform generous structure was relatively well suited to the economic and technological environment. But by 2019 it has become much less efficient relative to a "top-up" health insurance program where more basic public coverage can be supplemented by private health insurance. Our results are consistent with policies seen in many other developed countries, which provide a basic universal public insurance plan and where many citizens take advantage of the opportunity to pay extra for amenities and additional services.

Our model also helps to explain the puzzle of why the share of GDP devoted to health care has grown so much faster in the U.S. than in some other countries. The combination in the U.S. of an in-kind benefit reflecting the preferences of higher-income taxpayers coupled with lower tax rates (and commensurately lower excess burden of taxation) provided an ideal environment for rapid growth since 1965. In Europe, many countries had already elected to provide universal coverage, and with their high existing marginal tax rates, had limited scope for expanding benefits (Chandra and Skinner, 2012). By contrast, Medicare covered only a fraction of the population and marginal tax rates were relatively low, so the increase in spending on covered health care could, for a time, be absorbed without generating extremely high marginal tax rates. Expanding Medicare to a greater share of the population would only hasten the anticipated growth in tax rates necessary to fund it.

We explore the implications of an alternative "basic" form of public insurance that provides more restricted benefits with regulated prices and allows higher income households to top-up their coverage with privately financed plans. Under such a plan, lower income households would consume less health care than their higher income counterparts, and perhaps less care than they do now. This naturally raises concerns about equity. Many who support a uniform benefit structure point to the "right to health care" as a foundational rationale. It is worth noting, however, that a

uniform benefit is likely to result in a substantially higher share of income being devoted to health care (rather than, for example, food, education, or housing) than the typical lower-income household might choose. It is also likely to result in fiscal pressures over time that not only raise taxes but also crowd out spending on other public goods and transfer programs (Baicker, 2001). We demonstrate in our model that it is possible to offset some of the equity effects of the top-up redesign by appropriate redistribution of the taxpayer savings generated by scaling back the public benefit. Ultimately, the distributional implications of the plan redesign hinge on how the government uses this "Medicare dividend" – whether to reduce marginal tax rates on high-income households or increase social support for low-income households.

The tension between uniform (and thus "equitable") benefits and differential benefits extends to other public programs such as education. Universal higher education provided by the government and financed largely by taxpayers has been proposed by several 2020 presidential candidates and is provided in many European countries. In practice, because of budgetary pressure, many such "universal" higher education systems in other countries have converged on providing lower per-student expenditures, with wealthier students opting out into private universities.

This top-up structure is not only similar to that seen in some other countries, but is also related to other proposals discussed in the U.S. context, including value-based insurance design (Chernew et al., 2007) and premium support plans proposed by Emanuel and Fuchs (2005), Aaron and Reischauer (1995), and Ryan (2012).[21] Reinhardt (2012) suggested a three-tier system, with high-income households eligible for "the sky's the limit" insurance, largely unsubsidized by the government, middle-income households receiving a "reference pricing" insurance policy that reimburses only for the lowest-cost regional provider, and low-income enrollees receiving Medicaid-like public care under strict cost control rules.

Some of the current "Medicare-for-All" proposals simply extend the current uniform benefit to more people (or, in some cases, simultaneously increase the generosity).[22] Others suggest a

[21] While our model has focused primarily on "top-up" plans, other systems require higher-income households to step outside the insurance plan to pay entirely out-of-pocket, as is the case for, e.g., concierge medicine or treatments not covered under the National Health Service in England.

[22] To the extent that reforms such as alternative payment models improve the efficiency of the Medicare program, those improvements would also scale up with expansion. Expanding the Medicare Advantage program would also have different implications for the efficiency of delivery than expanding the traditional Medicare fee-for-service

hybrid "public option" in which Medicare (or even Medicaid) is provided as an option in addition to conventional private insurance – raising the questions of what payment rates the public plan would offer to providers, what network restrictions would be included, and how those features would affect competition between the private and public options. If the savings from choosing the cheapest option accrued to enrollees, such plans share some of the features of a top-up system.

Moving to a basic benefit-plus-top-up plan would of course pose both practical and political challenges because of the reliance on potentially controversial determinations of cost-effectiveness. Political pressures might, though, play out differently with different plans and different eligible populations. Following the birth of the National Health Service in Britain in 1948, cutbacks were quickly implemented in response to overly optimistic budget projections (Lane, 2012; pp. 191-93). A similar evolution could occur under a "Medicare for All" plan, in which budgetary pressures and monopsony power could lead to an erosion in reimbursements and a curtailment of benefits, creating a national program that more closely resembles "Medicaid for All" but with people free to top up or opt out into private plans.

As new technologies arrive with ever-larger price-tags, pressure will continue to mount on public budgets; and equality of access to care, rather than guaranteed access to a minimum level of care, will become increasingly costly. It is vital that policymakers consider how alternative program designs affect the overall wellbeing of households across the income distribution as they debate Medicare's future.

component. The important questions of efficient payment, cost-sharing, and network design within insurance plans are beyond the scope of this discussion.

References

Aaron, Henry J., and Robert D. Reischauer. 1995. "The Medicare reform debate: What is the next step?." *Health Affairs* 14, no. 4: 8-30.

Acemoglu, Daron, Amy Finkelstein, and Matthew J. Notowidigdo. 2013. "Income and Health Spending: Evidence from Oil Price Shocks." *The Review of Economics and Statistics* 95 no. 4:1079-1095.

ASPE. 2018. "Comparison of U.S. and International Prices for Top Medicare Part B Drugs by Total Expenditures." Office of the U.S. Assistant Secretary for Planning and Evaluation, Health and Human Services. https://aspe.hhs.gov/pdf-report/comparison-us-and-international-prices-top-medicare-part-b-drugs-total-expenditures

Baicker, Katherine. 2001. "Government Decision-Making and the Incidence of Federal Mandates." *Journal of Public Economics* 82, no. 2: 147-194.

Baicker, Katherine, Amitabh Chandra, and Jonathan Skinner. 2012. "Saving Money or Just Saving Lives? Improving the Productivity of US Health Care Spending." *Annual Review of Economics* 4 no. 1:33-56.

Baicker, Katherine, Mark Shepard, and Jonathan Skinner. 2013. "Public financing of the Medicare program will make its uniform structure increasingly costly to sustain." *Health Affairs* 32, no. 5: 882-890.

Bell, Felicitie and Michael Miller. 2005. "Life Tables for the United States Social Security Area 1900-2100." Social Security Administration, Actuarial Study No. 120. Available at https://www.ssa.gov/oact/NOTES/pdf_studies/study120.pdf

Blümel, Miriam, and Reinhard Busse. 2019. "The German Health Care System." International Commonwealth Fund. https://international.commonwealthfund.org/countries/germany/.

Bradley, Elizabeth H., and Lauren A. Taylor. 2013. *The American Health Care Paradox: Why spending more is getting us less*. New York, NY: Public Affairs.

Chandra, Amitabh, and Jonathan Skinner. 2012. "Technology growth and expenditure growth in health care." *Journal of Economic Literature* 50 no. 3:645-80.

Cheng, Terence C., Joan Costa-Font, and Nattavudh Powdthavee. 2018. "Do You Have to Win It to Fix It? A Longitudinal Study of Lottery Winners and Their Health-Care Demand." *American Journal of Health Economics* 4 no. 1:26-50.

Chernew, Michael. E., Allison B. Rosen, and A. Mark Fendrick. 2007. "Value-Based Insurance Design: By abandoning the archaic principle that all services must cost the same for all patients, we can move to a high-value health system." *Health Affairs* 26 no. Supplement 2:w195-w203.

Congressional Budget Office (CBO). 2019. "May 2019 Budget Baseline." https://www.cbo.gov/system/files/2019-06/55329-MBR.pdf.

Cooper, Zack, Stuart V Craig, Martin Gaynor, and John Van Reenen. 2018. "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured. "*The Quarterly Journal of Economics* 134 no. 1:51-107.

Currie, Janet, and Firouz Gahvari. 2008. "Transfers in Cash and In-Kind: Theory Meets the Data." *Journal of Economic Literature* 46 no. 2:333-83.

Cutler, David M., Allison B. Rosen, and Sandeep Vijan. 2006. "The Value of Medical Spending in the United States, 1960–2000." *New England Journal of Medicine* 355 no. 9: 920-927.

Emanuel, Ezekiel J, and Victor R. Fuchs. 2005. "Health care vouchers—a proposal for universal coverage." *New England Journal of Medicine* 352 no. 12:1255-60.

Glover, Lucinda. 2019. "The Australian Health Care System." International Commonwealth Fund. https://international.commonwealthfund.org/countries/australia/ .

Gottlieb, Joshua D., Adam Hale Shapiro, and Dunn Abe. 2018. "The Complexity of Billing and Paying for Physician Care." *Health Affairs* 37 no. 4:619-626.

Hall, Robert E., and Charles I. Jones. 2007. "The Value of Life and the Rise in Health Spending." *The Quarterly Journal of Economics* 122 no. 1:39-72.

Howard, David H., Peter B. Bach, Ernst R. Berndt, and Rena M. Conti. 2015. "Pricing in the Market for Anticancer Drugs." *Journal of Economic Perspectives* 29 no. 1:139-62.

International Federation of Health Plans. 2015. *International Federation of Health Plans 2015 Comparative Price Report: Variation in Medical and Hospital Prices by Country.*

Kamran, Sophia C., Jay O. Light, and Jason A. Efstathiou. 2019. "Proton versus photon-based radiation therapy for prostate cancer: emerging evidence and considerations in the era of value-based cancer care." *Prostate Cancer and Prostatic Diseases:*1-13.

Kullgren, Jeffrey T., Alison A. Galbraith, Virginia L. Hinrichsen, Irina Miroshnik, Robert B. Penfold, Meredith B. Rosenthal, Bruce E. Landon, and Tracy A. Lieu. 2010. "Health Care Use and Decision Making Among Lower-Income Families in High-Deductible Health Plans." *Archives of Internal Medicine* 170 no. 21:1918-1925.

Jaffe, Sonia, and Mark Shepard. 2018. "Price-Linked Subsidies and Imperfect Competition in Health Insurance." NBER Working Paper 23104.

Lane, Joan 2012. *A Social History of Medicine Health, Healing and Disease in England, 1750-1950*: Routledge.

Maeda, Jared Lane, and Lyle Nelson. 2017. "An analysis of private-sector prices for hospital admissions." Working Paper 2017-02, Congressional Budget Office.

Mertens, Karel, and José Luis Montiel Olea. 2018. "Marginal tax rates and income: New time series evidence." *The Quarterly Journal of Economics* 133, no. 4: 1803-1884.

National Health Service. 2018. Clinical Commissioning Policy: Proton Beam Therapy for Cancer of the Prostate. National Health Service England: 16020/P.

Papanicolas, Irene, Liana R. Woskie, E. Duncan Orlander, John Orav, and and Ashish K. Jha. 2019. "The Relationship Between Health Spending and Social Spending In High-Income Countries: How Does the US Compare?" *Health Affairs* 38 no. 9:1567-1575.

Piketty, Thomas, and Emmanuel Saez. 2014. "Inequality in the Long Run." *Science* 344 no. 6186: 838-843.

International Federation of Health Plans. 2015. *International Federation of Health Plans 2015 Comparative Price Report: Variation in Medical and Hospital Prices by Country*: International Federation of Health Plans.

Prostate Cancer Canada. 2019. "Proton Therapy." Prostate Cancer Canada. https://www.prostatecancer.ca/Prostate-Cancer/Treatment/Treatment-Options/Emerging-Treatments/Proton-Therapy.

Reinhardt, U. 2012. "A Fork in the Road for Health Care." *The New York Times.* May 25, 2012.

Ryan, Paul. 2012. *The Path to Prosperity: A Blueprint for American Renewal.* Washington, DC: Fiscal Year 2013 Budget Resolution.

Saez, Emmanuel, Joel Slemrod, and Seth H. Giertz. 2012. "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *Journal of Economic Literature* 50, no. 1: 3-50.

Schroeck, Florian Rudolf, Bruce L. Jacobs, Sam B. Bhayani, Paul L. Nguyen, David Penson, and and Jim Hu. 2017. "Cost of new technologies in prostate cancer treatment: systematic review of costs and cost effectiveness of robotic-assisted laparoscopic prostatectomy, intensity-modulated radiotherapy, and proton beam therapy." *European Urology* 72 no. 5:712-735.

Shepard, Mark. 2016. "Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange." No. w22600. National Bureau of Economic Research.

Sturny, Isabelle. 2019. "The Swiss Health Care System." International Commonwealth Fund. https://international.commonwealthfund.org/countries/switzerland/ .

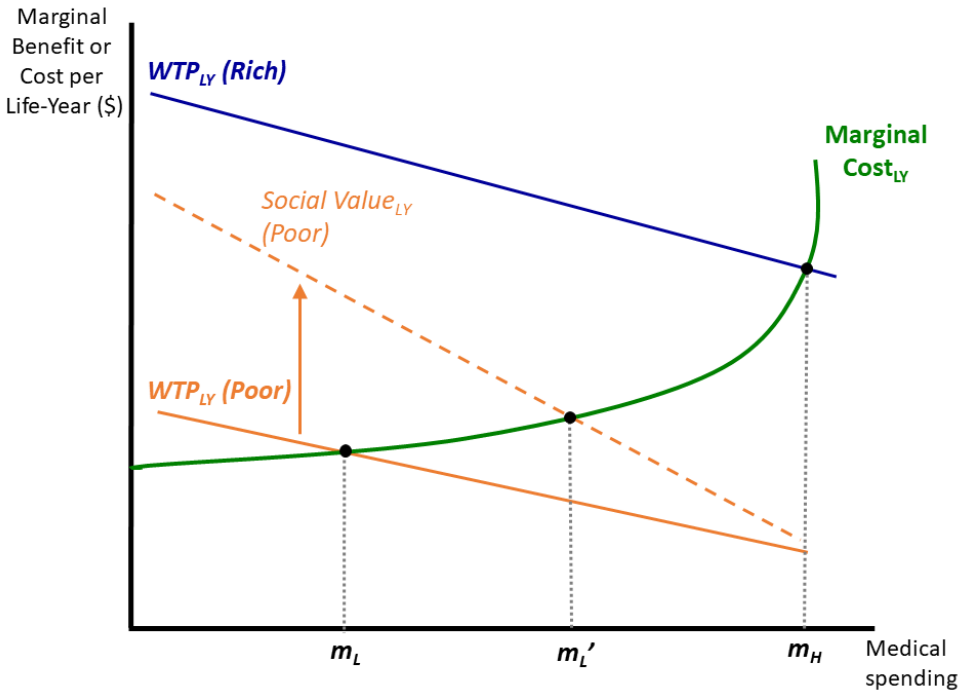
Tebaldi, Pietro. 2017 "Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design under the ACA." Becker Friedman Institute for Research in Economics Working paper 2017-05.

Thorlby, Ruth, and Sandeepa Arora. 2019. "The English Health Care System." International Commonwealth Fund. https://international.commonwealthfund.org/countries/england/

Figures

Figure 1: Graphical Model of Private and Public Medical Spending

Panel A: Private Medical Spending Choices



Panel B: Uniform Benefits

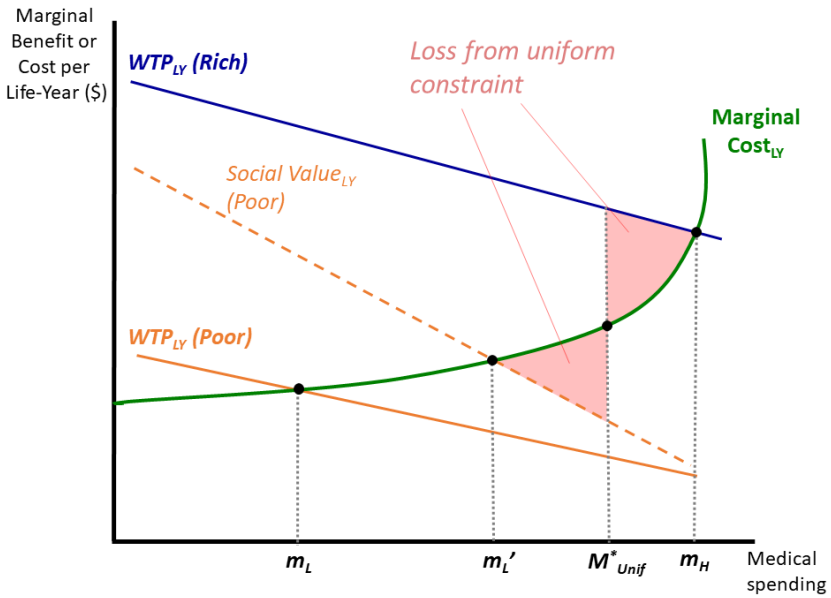


Figure 2: Top-Up vs. Uniform Benefits

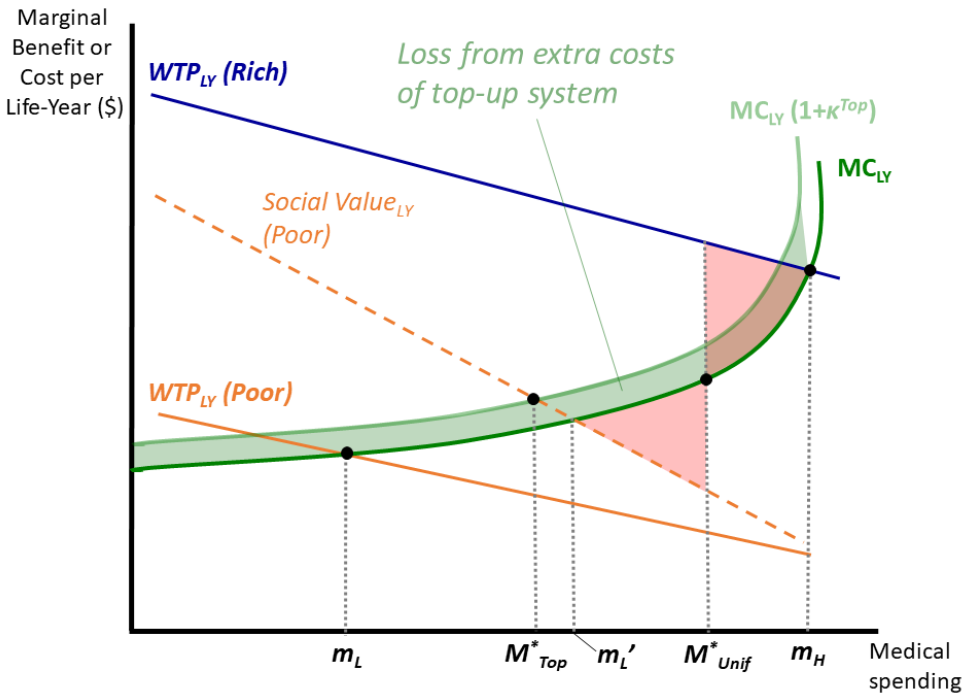
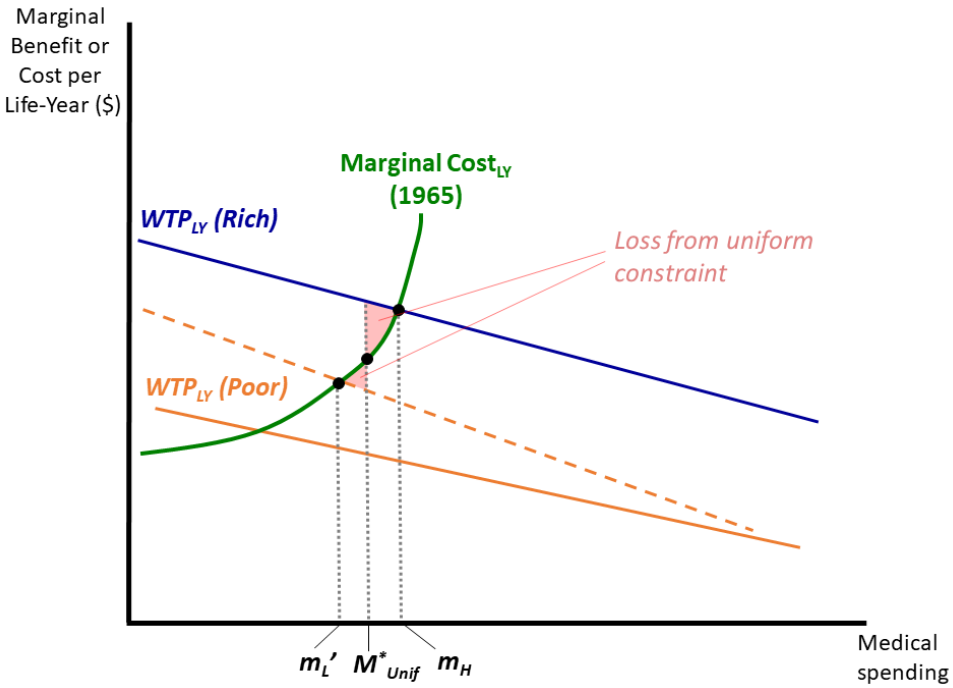
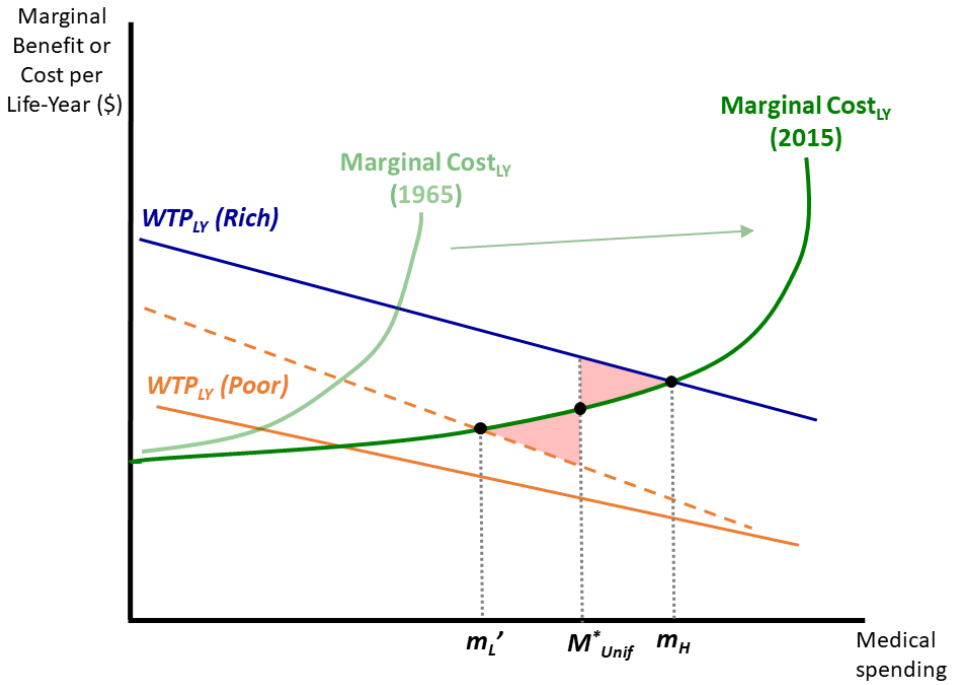


Figure 3: Rising Costs of Uniform Health Insurance Benefits

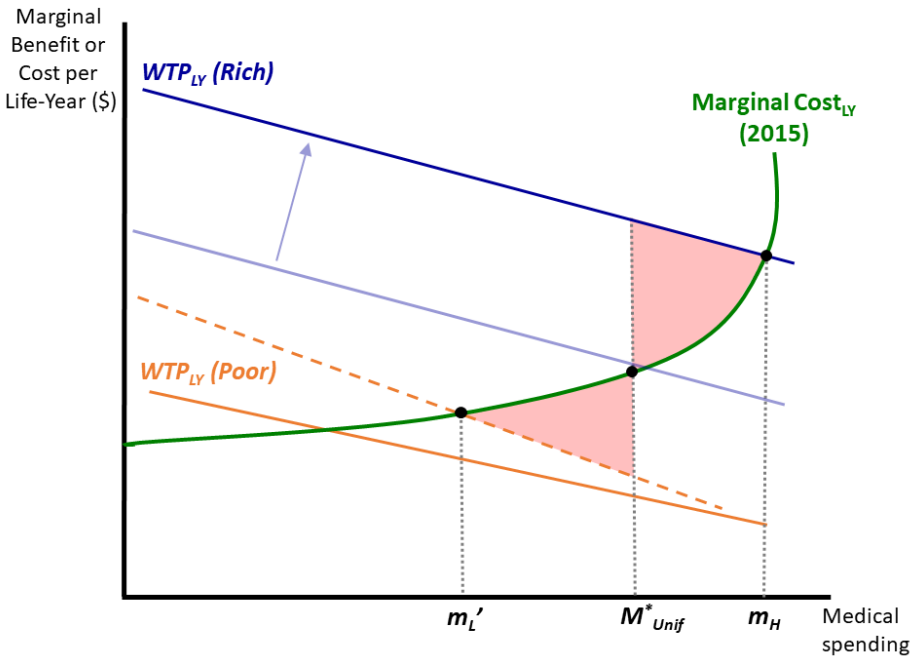
Panel A: Baseline (Medicare in 1965)



Panel B: Improving Medical Technologies



Panel C: Rising Income Inequality



Panel D: Rising Deadweight Loss (DWL) of Taxation

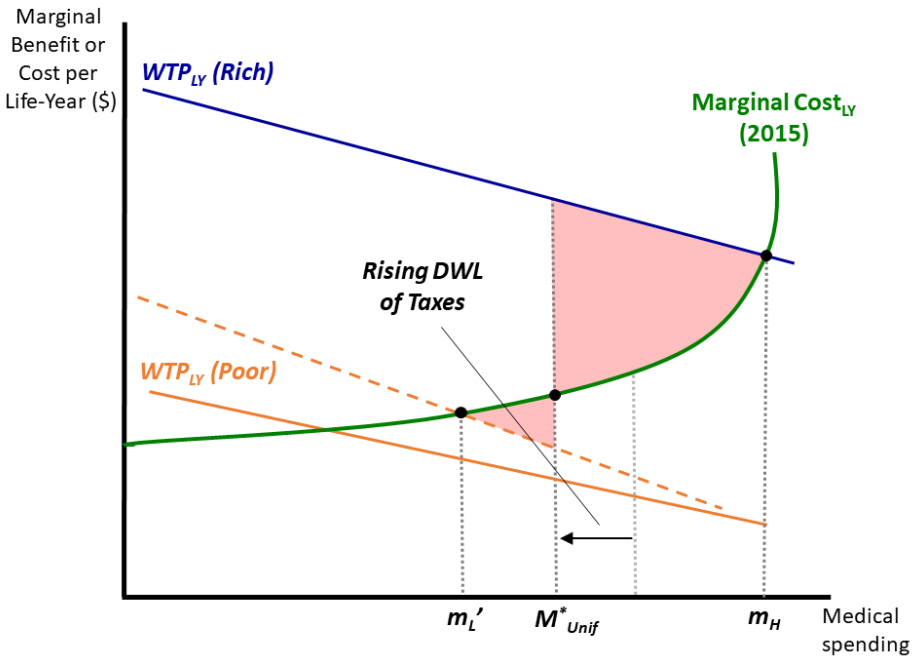


Figure 4: Simulations: Widening Inequality in Private Medical Spending

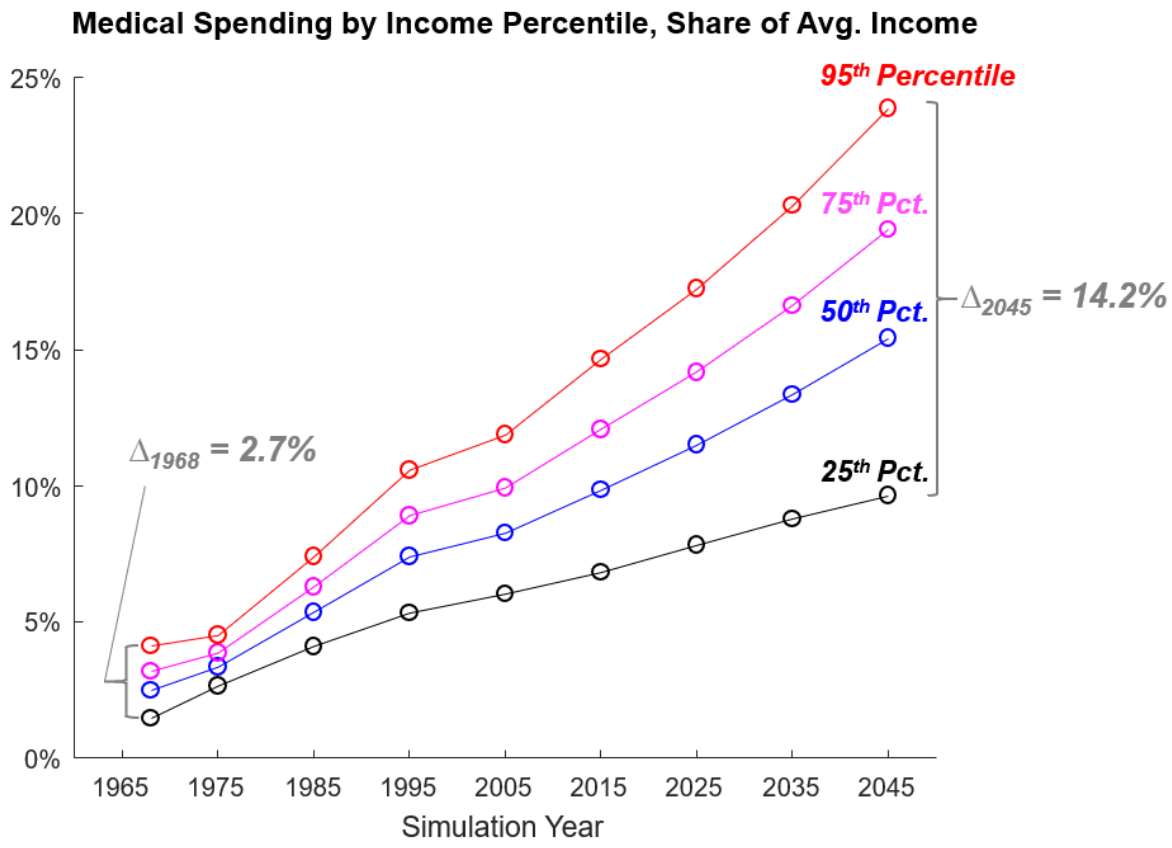


Figure 5: Simulations: Public Medical Benefit Level

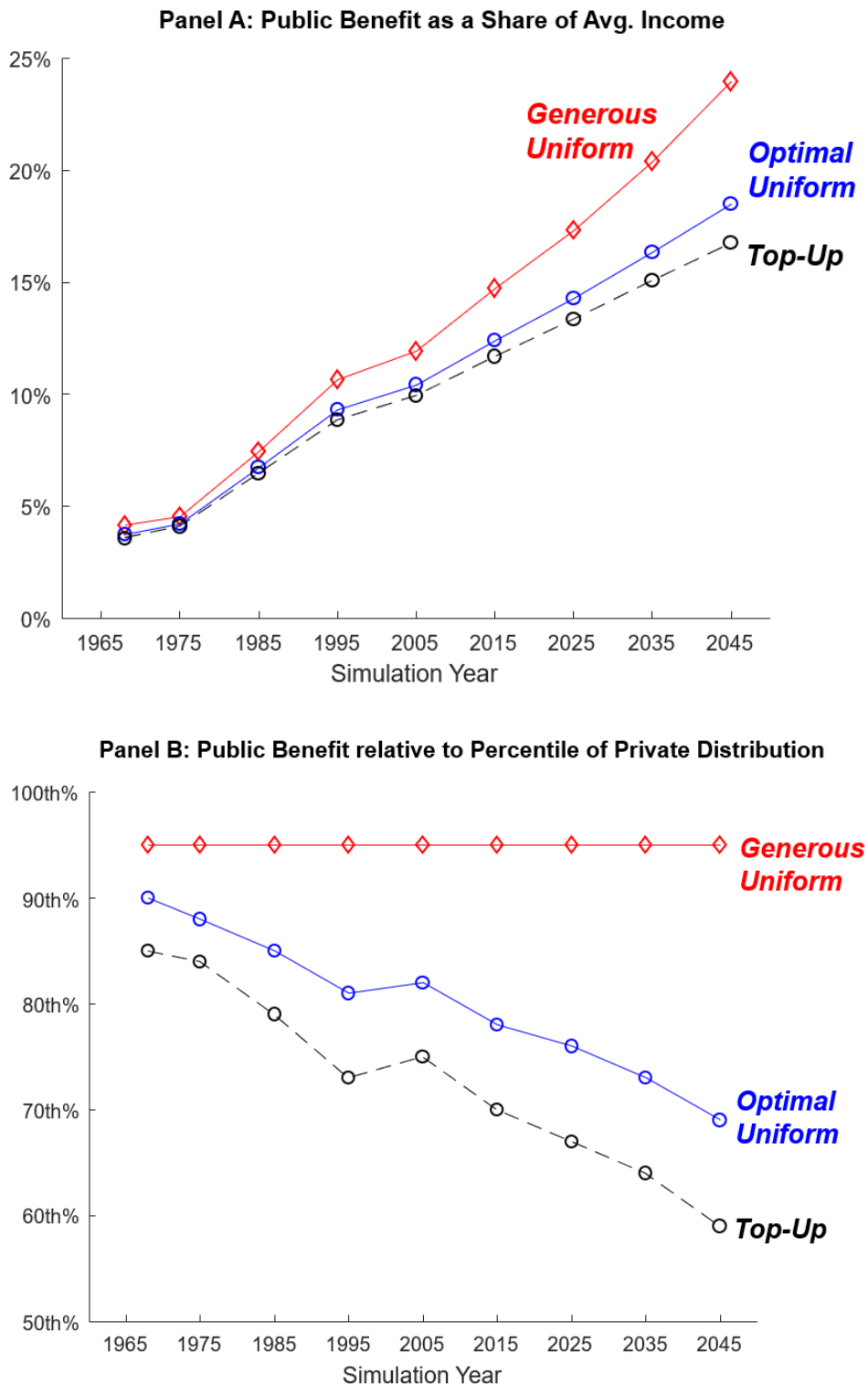


Figure 6: Simulations: Top Marginal Tax Rates

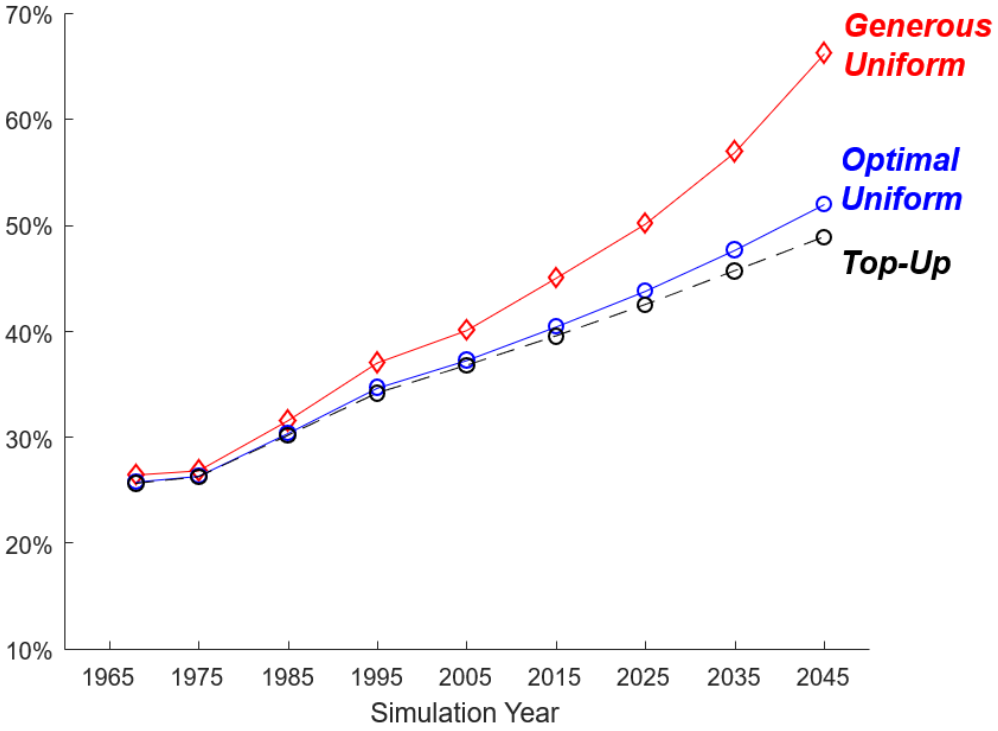
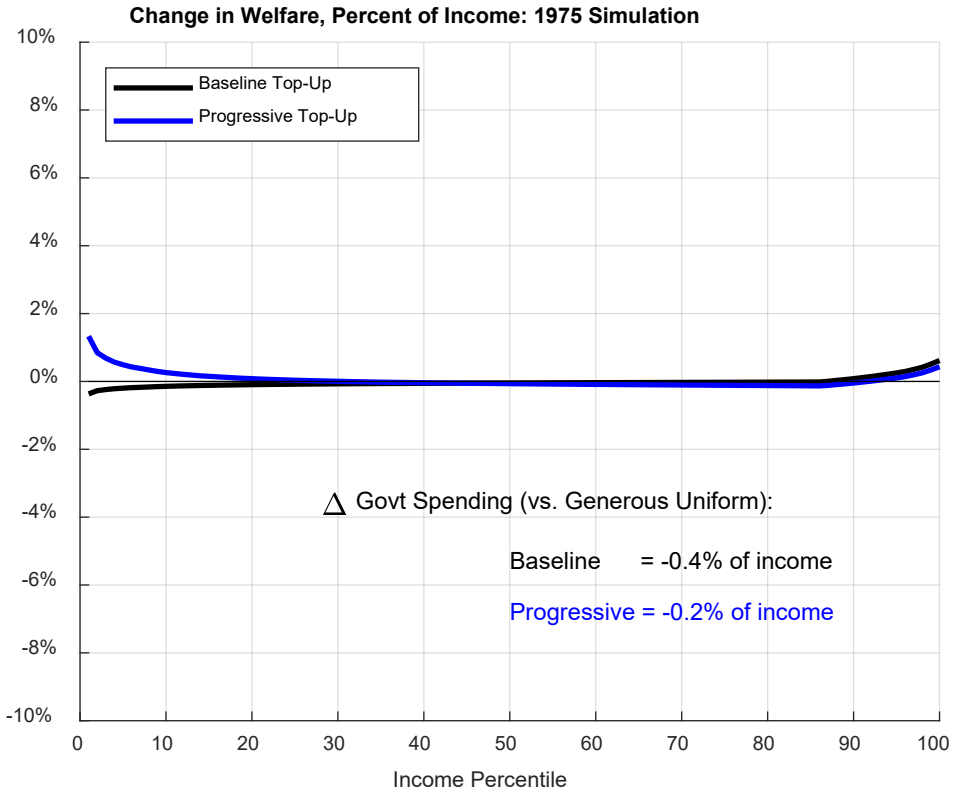
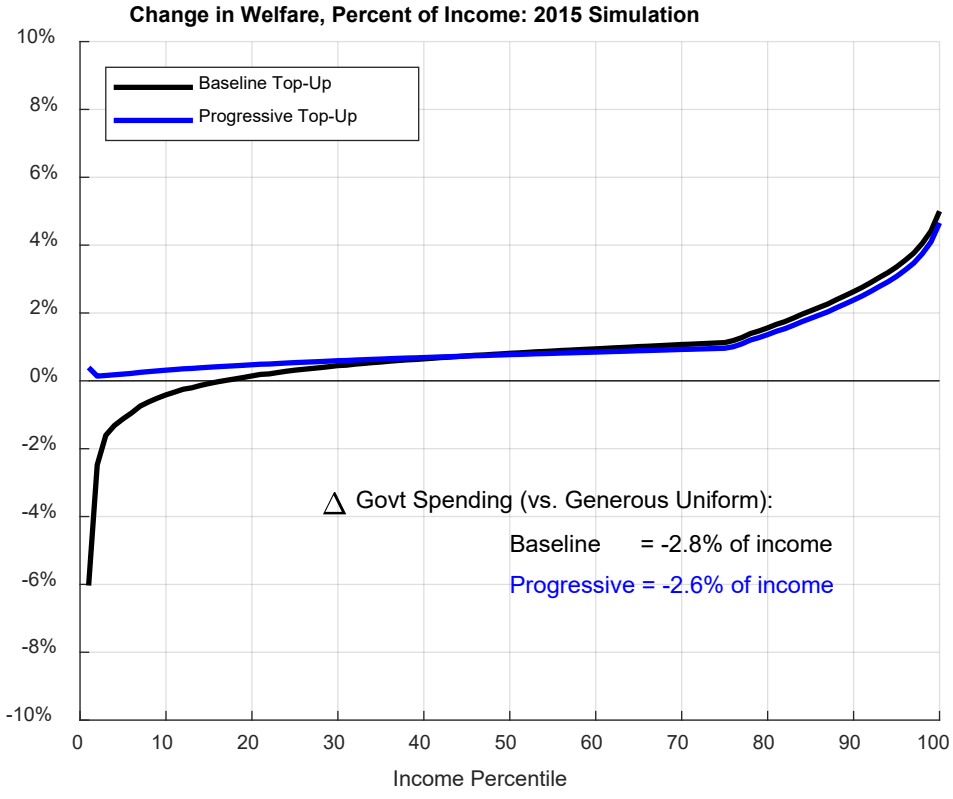
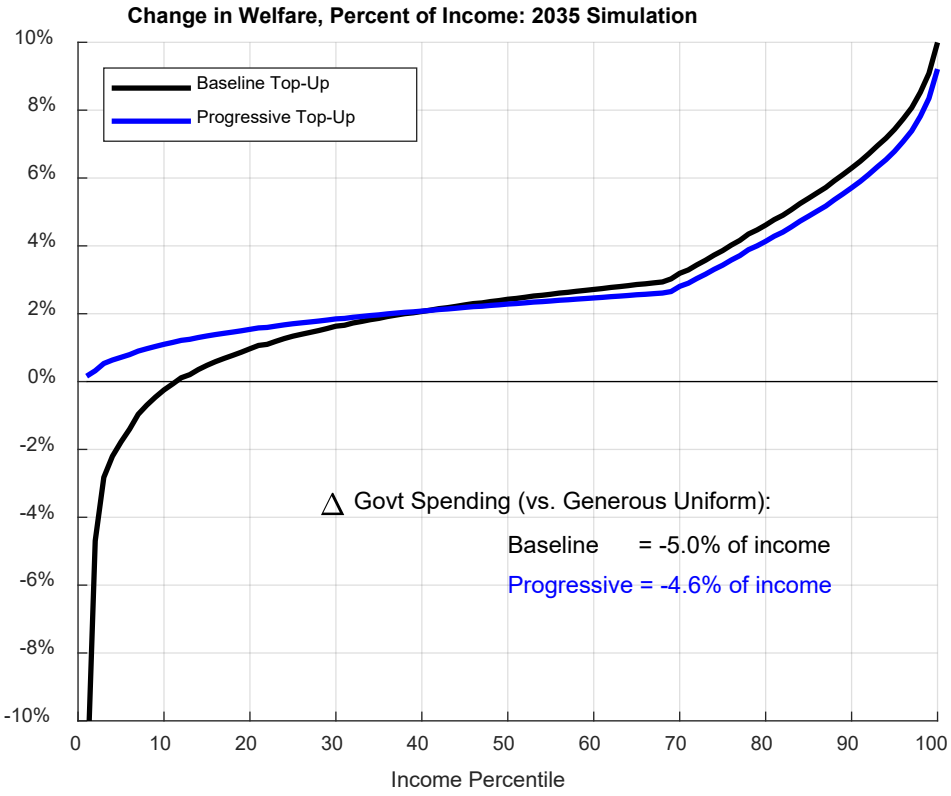


Figure 7: Distributional Welfare Impacts of Shift to Top-Up Benefits






Change in Welfare, Percent of Income: 2035 Simulation
Baseline Top-Up
Progressive Top-Up
△ Govt Spending (vs. Generous Uniform):
Baseline = -5.0% of income
Progressive = -4.6% of income
10%
8%
6%
4%
2%
0%
-2%
-4%
-6%
-8%
-10%
0
10
20
30
40
50
60
70
80
90
100
Income Percentile

Appendix A: Simulation Model Details

This appendix presents the details of the specification and calibration of our simulation model. The model takes the form of standard public economics two-level optimization, with the government choosing economic policy and individuals making choices given the government policy. We discuss these two parts in reverse order. We start by recapping the model setup for each part and then discuss specification and calibration details.

Individual Problem

As shown in equation (1) in the text, the individual (*i*) optimization problem is:

$$\max_{c_i, m_i} \lambda\left(M + m_i\right) \cdot u\left(c_i\right)$$
$$\text{s.t.} \quad c_i + m_i = y_i - T\left(y_i\right) + R$$

where M is the public medical benefit and m_i is extra individual health spending, $\lambda(M + m_i)$ is longevity as a function of total medical spending $M + m_i$ (i.e., the health production function), $u(c)$ is flow utility of consumption, and y_i is individual income, $T(y_i)$ is income taxes owed, and R is a government cash transfer.

The individual also faces a constraint on extra individual health spending that depends on the public insurance system. We consider two possibilities:

1. **Uniform benefit:** $m_i = 0$ *(individuals cannot top-up the public benefit)*
2. **Top-up benefit:** $m_i \geq 0$ *(individuals can top-up above M with private funds)*

Note that this setup also allows for a purely private health system, which is equivalent to a top-up benefit with $M = 0$.

Function Specifications: We make the following specification choices for $u(.)$, $T(.)$, and $\lambda(.)$:

- Flow utility: $u(c) = b + \log\left(c\right)$, where b is a constant set so that the implied value of a life-year is \$100,000 at $c = \$30{,}000$. We have also experimented with other constant relative risk aversion specifications (e.g., CRRA = 2) and found little change in the main results.

- Taxes: We model taxes as a scaled version of a baseline progressive schedule $T_0(y_i)$, so that $T(y_i) = \tau \cdot T_0(y_i)$ where τ is the scale factor that the government sets to balance its budget. The baseline tax schedule $T_0(y)$ is set to be a smoothed version of the U.S. income, payroll, plus state tax schedules. This function smooths over jumps in marginal tax rates (which simplifies computation) and ensures globally increasing marginal tax rates, to avoid problems associated with non-convex budget sets. We first fit a polynomial function to the marginal tax rate schedule (with arguments of $\log(y + 2000)$ to improve the fit) then integrate up to calculate the smoothed taxes owed schedule. Appendix Figure A.1 shows the resulting fit.

- Health production function: The health production function captures the map between medical spending (m) and expected longevity/health. To specify this, we start by specifying a function for an annual survival rate $\sigma(m)$ (one minus the mortality rate):

$$\sigma(m) = s_0 + (1 - \exp(-\alpha \cdot m)) \cdot (s_{\max} - s_0)$$

This function implies that an individual with no health care (m = \$0) has a survival rate of s_0, and as $m \to \infty$ the survival rate approaches $s_{\max}$. The marginal returns to a given level of medical spending (i.e., the function concavity) are determined by α, with a larger α implying more concavity or more steeply diminishing returns at low levels of spending. Total life expectancy at age 65 equal the inverse of the mortality rate, or:

$$\lambda(m) = \frac{1}{1 - \sigma(m)}$$

which is the life expectancy implied by a constant mortality rate of $1 - \sigma(m)$. We fix $s_0 = 0.75$ in all years but estimate $s_{\max}$ and α separately to match spending and life expectancy moments in each simulation year. To match the growing spending and improving longevity, we estimate that α declines (from 0.0012 in 1968 to 0.00016 in 2045) and $s_{\max}$ increases (from 0.82 in 1968 to 0.94 in 2045).

Government Policy Problem

The government makes the following policy choices. It chooses the tax scalar (τ), cash transfer (R), medical benefit (M), and design of medical benefits to maximize the social welfare function:

$$SW = \sum_i \left[\lambda(m_i) u(c_i) + e(m_i) \right]$$

subject to public budget constraint:

$$\sum_i \left(R + M \cdot (1+\kappa) + E \right) = \sum_i \tau \cdot T_0(y_i) \cdot (1 - \chi(\tau))$$

and where $\{m_i, c_i\}$ are set by individual choices from the problem in (1) and κ captures the excess administrative costs of providing in-kind medical benefits as opposed to cash. The functions and parameters that need to be specified here are:

- Egalitarian externality ($e(m)$): The egalitarian externality determines the extra social value of medical spending that motivates in-kind medical benefits (as opposed to purely cash transfers). We specify it as the additional social value of living an extra year, valued at flow utility of consuming the year's median income, $u(\bar{y}_t)$. Mathematically:

  $$e(m) = \lambda(m) \cdot u(\bar{y}_t)$$

  By using a fixed $u(\bar{y}_t)$ for all individuals, it effectively places more weight on improving the health of the poor relative to their private value of health.

- Deadweight loss of taxes: To capture these losses, we do not model a full labor supply problem but instead specify that levying taxes that reduce individuals' consumption by \$1 only raises tax revenue of $\$(1 - \chi(\tau))$, where $\chi(\tau)$ captures revenue leakage due to the excess burden of taxes. As taxes rise, $\chi(\tau)$ will increase due to the rising marginal excess burden of higher tax rates. We specify the excess burden term as $\chi(\tau) = \varepsilon_{TI} \cdot \overline{MTR}(\tau)$, where ε_{TI} is the elasticity of taxable income (assumed to be 0.5) and $\overline{MTR}(\tau)$ is the average marginal tax rate for tax scalar τ. This function for $\chi(\tau)$ is derived from a

calculation of average revenue leakage (starting from an initial tax of 0) assuming a constant elasticity of taxable income. To see this, note that by definition:

$$d\log\left(Rev\right) = \varepsilon_{TI} \cdot d\log\left(1 - MTR\right)$$
$$= \varepsilon_{TI} \cdot \frac{-dMTR}{1 - MTR}$$

Now consider a particular income group *g* and consider a tax increase from *MTR* = 0 up to the rate implied by tax scalar τ, which is $MTR_{0,g} \cdot \tau$ where $MTR_{0,g}$ is the group's baseline marginal tax rate given the baseline tax schedule. We then adapt the formula for $d\log(Rev)$ to approximate this discrete tax increase (starting from an initial *MTR* = 0):

$$\%\Delta Rev_g \approx \varepsilon_{TI} \cdot \frac{\tau \cdot MTR_{0,g} - 0}{1 - 0} = \varepsilon_{TI} \cdot \tau \cdot MTR_{0,g}$$

We then calculate an income-weighted average of this $\%\Delta Rev_g$ term across income groups to generate $\chi\left(\tau\right) = \overline{\%\Delta Rev} = \varepsilon_{TI} \cdot \tau \cdot \overline{MTR_{0,g}} = \varepsilon_{TI} \cdot \overline{MTR}\left(\tau\right)$. This captures a measure of the *average* percent change in revenue (or average "revenue leakage") from the tax increase from zero. Note that this differs from the more familiar formula for the *marginal* revenue leakage of a tax increase (see Saez, Slemrod, and Giertz, 2012).

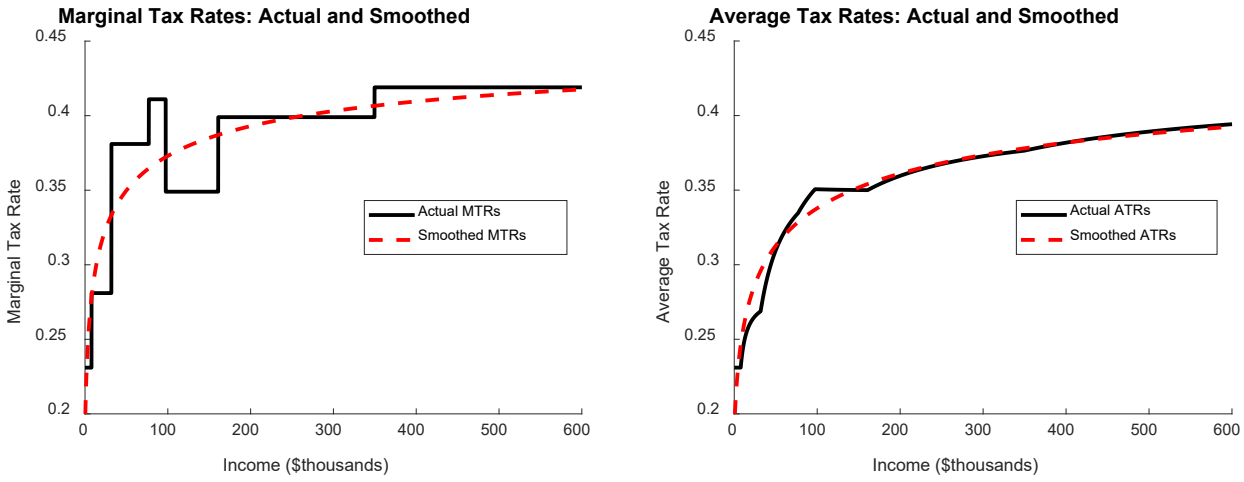
- Administrative costs (κ): There is little evidence from which to draw on in specifying the costs due to administration. In practice, only the difference in κ between the uniform and top-up settings is relevant for our analysis. We therefore normalize $\kappa = 0$ for the uniform systems (generous and optimal). We set $\kappa = 2\%$ for top-up benefits in order to add a small additional value of administrative costs ensures that the top-up policy need not always be socially preferable. In practice, this results in the optimal uniform benefits being socially preferred over top-up in the earlier years of the model and top-up benefits being preferred in later years. Both top-up and optimal uniform yield higher welfare than the generous uniform specification in all years.

Calibration to Medical Spending and Life Expectancy Data

We flexibly calibrate the parameters $s_{\max}$ and α of the health production function for each simulation year to match two moments: (1) the historical or projected Medicare spending per beneficiary, drawing on projections from the Medicare trustees, and (2) historical or projected average life expectancy at age 65 from the Social Security Administration (Bell and Miller, 2005). We match these moments under the policy of generous uniform medical benefits (i.e., M set at what the 95th income percentile would choose), since this is intended to proxy for the Medicare policy in place to generate the historical data.

By flexibly calibrating $\lambda(m)$ in each simulation year, we capture the way that technology improves to both extend life (and, more generally, quality of life) and to induce people to spend more on medical care. Because the estimation is just-identified, we match all of the moments exactly. Appendix Table A.1 lists the moments and parameter estimates for each simulation year.

Appendix Figure A.1: Smoothed Tax Rate Functions for Model



Appendix Table A.1: Moments and Parameter Estimates

Simulation Year	Moments		Parameter Estimates	
	Medicare Spending per Elderly Person (in 2015 $)	Life Expectancy at Age 65	Smax	Alpha
1968	$2,637	70.5	0.821	1.16E-03
1975	$3,040	72.7	0.871	1.52E-03
1985	$5,268	74.7	0.898	9.27E-04
1995	$7,845	75.9	0.909	6.13E-04
2005	$10,182	77.2	0.919	4.97E-04
2015	$12,741	78.2	0.925	4.00E-04
2025	$17,345	79.1	0.930	2.94E-04
2035	$23,634	80.1	0.935	2.15E-04
2045	$32,182	80.9	0.938	1.57E-04