Aggregate Effects of Public Health Insurance Expansion: The Role of Delayed Medical Care

Mitchell VanVuren (Yale)

October 2023

Please click here for the latest version

Abstract

Evidence indicates many U.S. adults postpone medical care until they qualify for Medicare at age 65. This paper studies the aggregate consequences of expanding public health insurance access (e.g. Medicaid), accounting for reductions in delayed care, using a two-asset overlapping generations model with health investment and endogenous mortality. Reducing delayed care results in long-run savings as earlier treatment is less costly but increases total Medicare outlays as lower mortality leads to a larger senior population. I estimate the model to match quasi-experimental evidence on the extent of delayed care in U.S. adults and the impacts of Medicaid expansion on mortality and find that the savings channel dominates and is roughly twice as large as the mortality channel. A policy that gradually phases in insurance coverage for the elderly is more cost-effective, reducing mortality by substantially more per dollar spent, but is more limited in scope.

I thank Titan Alon, Mariacristina De Nardi, Roozbeh Hosseini, Minki Kim, Karen Kopecky, David Lagakos, Vegard Nygaard, Valerie Ramey, Kai Zhao, and various seminar participants at the SED annual meeting in Cartagena, the Federal Reserve Bank of Minneapolis, and UCSD for their helpful comments. Any errors, of course, remain my own. Mitchell VanVuren: Yale Research Initiative on Innovation and Scale (Y-RISE) email: mitchell.vanvuren@yale.edu

1. Introduction

A substantial body of evidence suggests that a large fraction of U.S. adults delay medical care until after age 65 when they become eligible for Medicare. For example, Card, Dobkin & Maestas (2008) document that many usage of medical procedures – including doctor's visits, heart surgery, and gall bladder removals – jumps discretely at age 65. Furthermore, McWilliams et al. (2003) show that the use of testing services like cholesterol testing, mammography, and prostate examination rise substantially for uninsured individuals right after they turn 65, and Patel et al. (2021) show that cancer diagnoses rise substantially at age 65, particularly for early-stage cancers.

This delayed medical care carries potentially large financial costs. Nearly 80 percent of adults approaching Medicare eligibility (ages 55-64) have been diagnosed with at least one chronic health condition and 37 percent have been diagnosed with at least two (CDC 2009), and many medical studies (e.g. Gehi et al. 2007, Herkert et al. 2019, Fukuda & Mizobe 2017) show that delaying treatment of these conditions risks both the individual's life and can lead to higher eventual treatment costs as the disease progresses. As an illustrative example, mild cases of coronary artery disease (i.e. the build-up of plaque in one's arteries) can be treated at relatively low cost through medications that help prevent or reduce the blockage of arteries. Atorvastatin, the generic version of popular anti-cholesterol medication Lipitor, costs roughly $20 per month. However, if a mild case worsens due to delayed care, it can require surgical treatment, such as bypass surgery, which carries an average cost of $169,000 (Benjamin et al. 2018). Some portion of those who delay care do so with deadly consequence. Miller, Johnson & Wherry (2021) show that receiving Medicaid coverage reduced all-cause mortality of low-income individuals aged 55-64 by 10 percent, consistent with the notion that these individuals are delaying important medical treatment when uninsured.

In this paper, I study the aggregate consequences of delayed medical care for public health insurance expansion in the spirit of the macro literature on health (e.g. De Nardi, French & Jones 2016). I focus on two channels. First, expanding public health insurance can reduce delayed care and, since early treatment tends to be less expensive than later treatment, result in long-run cost savings. Second, expanding public insurance can raise the total number of people over age 65 since earlier care tends to reduce mortality, but this raises long-run costs. Although both effects are ex-ante welfare increasing, they have opposing impacts on the total cost of expansion.

To study these channels, I construct a heterogeneous-agent overlapping generations general equilibrium model featuring health investment and endogenous mortality. Following much of the literature on macroeconomics and health, individuals build and maintain health capital through medical spending each period (see Fang & Krueger 2021, for an overview). Health capital reduces an individual's mortality risk, as in Ozkan (2014), as well as their chance of experiencing a costly health emergency. Individuals face a choice between purchasing health insurance or not which leaves some low-income individuals uninsured until they receive Medicare at age 65. For uninsured individuals approaching age 65, the optimal strategy is indeed one of delaying healthcare spending; they treat their health as an asset and substitute consumption for medical care, running down their health capital. After turning 65 and gaining health insurance, these individuals compensate for their period of low spending through higher use of medical care; however, some individuals die or end up needing more expensive medical care as a result of delaying care.

I estimate the model to match, among other things, two key quasi-experimental studies from the health literature, each of which provides discipline to one of the channels. First, Card, Dobkin & Maestas (2008) use administrative data from hospitals and a regression discontinuity design to show that individuals receive 54 percent more medical procedures immediately after turning 65. This disciplines the intertemporal elasticity of substitution of healthcare spending in the health production function (if this elasticity is high, individuals will be very willing to delay care as it carries little cost) which determines the extent to which reductions in delayed care lead to overall cost savings. Second, Miller et al. (2021) use a diff-in-diff framework to show that expansion of Medicaid (the US's public insurance program for poor individuals) reduced mortality for low-income near-old (ages 55-64) individuals by 10 percent. This pins down the overall productivity of the health production function (if productivity is high, gaining insurance leads to a large decline in mortality) and determines the magnitude of the cost increase due to lower mortality.

The estimated model accurately reproduces many untargeted moments related to health and the expansion of public health insurance. Most importantly, it closely matches the strong left-skew exhibited by the distribution of health — many individuals are in perfect or close-to-perfect health while a few individuals are extremely sick — while only the average level of health is included as an estimation target. The model also replicates the longer-run dynamics of the decline in mortality following Medicaid expansion, predicting a decline of 14 percent three years after the initial expansion compared to a measured decline of about 12 percent, despite only the initial decline being included in model estimation. Finally, the model closely reproduces an array of untargeted aggregate statistics such as the overall cost of public insurance expansion under the 2014 ACA, the percentage of the population covered by public insurance before and after expansion, and the difference in average life expectancy between those who do and do not qualify for Medicaid.

I use the estimated model to evaluate the impact of an expansion of public health insurance similar to the expansion of Medicaid under the 2014 ACA, funded by an increase in distortionary taxes. Expansion modestly reduces delayed care, particularly for individuals between ages 60 and 64 who are approaching the Medicare qualification threshold of 65. Average annual medical investment spending increases by 7 percent and annual mortality declines by about 0.3 percentage points for these individuals. Consistent with a reduction in delayed care, individual medical spending after age 65 falls by about 4 percent.

Overall, I find that the cost-savings channel stemming from this reduction in delayed care outweighs the increase in costs due to lower mortality. For every \$100 spent on the expansion, total Medicare outlays fall by \$8. In other words, 8 percent of the cost of expansion is offset by lower Medicare outlays. In order to quantify each channel separately, I consider an alternative model in which mortality is exogenous and, consequently, the increase in costs due to the decline in mortality is zero leaving only the cost savings channel. In this model, Medicare outlays decline by \$15 for every \$100 spent on expansion when prices are fixed and \$14 per \$100 when prices are allowed to adjust. I interpret this to mean that the decline of \$8 (per \$100) in the baseline model is the net result of a \$15 decline due to less delayed care, a \$6 increase due to lower mortality, and a \$1 increase due to price adjustments.

Despite driving a substantial increase in insurance coverage, the expansion of public health insurance results in only a modest decline in delayed care; the jump in spending at age 65 falls by about 7 percentage points (from 57 percent to 50 percent). Examining how delayed care varies by income reveals the underlying reason; although low-income individuals exhibit more substantial delay of care due to higher rates of uninsurance and a higher likelihood of being credit constrained, even high-income individuals (who are not impacted by the expansion) substantially delay care. This behavior is a result of that fact that the transition to Medicare represents an improvement in insurance coverage (at least in terms of the marginal cost of care faced by the individual) even for the insured and highlights the fact that it is the sudden decline in the marginal cost of healthcare, rather than credit constraints, that mostly drive delay.

Prompted by this fact, I examine a policy that slowly phases in public insurance for the elderly (i.e. Medicare) in order to eliminate this sudden jump in the marginal cost of healthcare. The phase-in policy targets individuals between ages 61 and 64 and is structured such that the gap between an individual's marginal cost of healthcare (under their private insurance plan) and their marginal cost of healthcare at age 65 (under

Medicare) shrinks by 20 percent per year — that is, 20 percent of the gap is closed at age 61, 40 percent at age 62, and so on. This policy completely eliminates delayed care in the sense that the jump in healthcare at age 65 is reduced to 0 percent. Similar to the Medicaid expansion, this policy increases healthcare spending for those between the ages of 61 and 64, reduces spending for those 65 and older, and reduces mortality.

Because it so effectively reduces delayed care, the phase-in policy results in a substantial decline in Medicare outlays. For every $100 spent on the policy, Medicare spending falls by $64, offsetting almost two-thirds of the cost of the policy and demonstrating that there are significant cost-reductions that can be realized by eliminating delayed care. Because of this large offset, the policy is substantially more cost-effective than the expansion policy, reducing mortality between the ages of 60 and 70 by 1 percentage point per $12 billion spent (compared to $18 billion in the expansion policy). Still, the small size of the phase-in policy (compared to expansion) with no natural way to increase its scope suggests that the two policies should be considered complements rather than substitutes.

Related Literature: This paper is inspired by and builds on a growing macroeconomic literature evaluating the role of public health insurance using macroeconomic models. De Nardi, French & Jones (2016) evaluate Medicaid in the context of late-in-life insurance and find that it is approximately the correct size. Aizawa & Fu (2020) examine the interaction between risk-pool cross subsidization and Medicaid expansion and find that expansion leads to higher welfare gains. Kopecky & Koreshkova (2014) examine the impact of health insurance and late-in-life medical expenses on wealth accumulation. Pashchenko & Porapakkarm (2013) evaluate whether the welfare gains from expansion come from primarily regulatory changes or primarily redistribution and find that the welfare gains overwhelming come from the latter. Jung & Tran (2016) examine this same question in a more complex model with endogenous health expenditure and find a similar answer. To measure health, I leverage the approach of Hosseini et al. (2021) by constructing frailty indices.

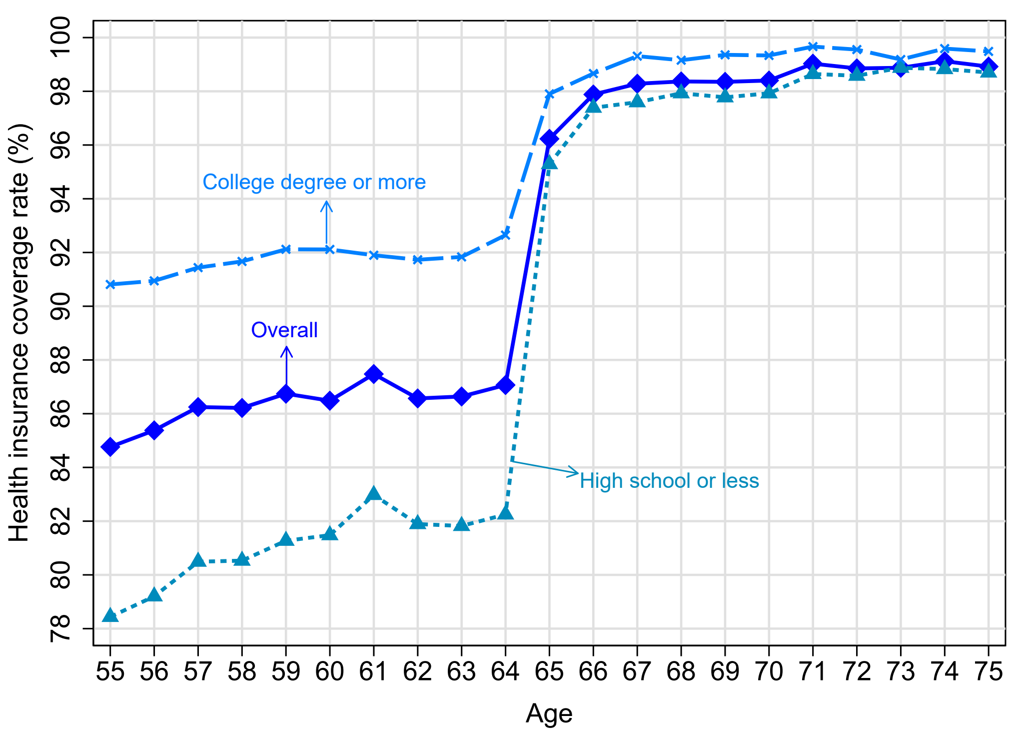
This paper also contributes to the literature on macroeconomics and insurance such as Kaplan & Violante (2010). I add to a large and growing literature on self-insurance in two-asset heterogeneous agent models such as Kaplan, Moll & Violante (2018) since health functions as an asset in the model. In a similar vein, health in the model can also be thought of as a durable good as in McKay & Wieland (2019).

This paper is most closely related to the work of Ozkan (2014) which estimates a macroeconomic model of health spending and argues that shorter optimal lifespans for poorer individuals cause these individuals to under-spend on preventative care early in life, face a more costly distribution of late-in-life health shocks, and spend more on healthcare overall. My paper, instead of focusing on early-in-life preventative care, focuses on the incentives to delay the treatment of already-developed conditions induced by the age threshold of Medicare.

Finally, this paper also contributes to the broader literature on health and healthcare spending in macroeconomic models. De Nardi, French & Jones (2010) examine the role that late-in-life medical expenses play in individuals' optimal savings behavior. Cole, Kim & Krueger (2019) estimate optimal insurance policy in a model where health and labor market risks are intertwined and health insurance induces a moral hazard inefficiency.

2. Some Facts on Healthcare Near Age 65

Figure 1: Health Insurance Coverage by Age and Education

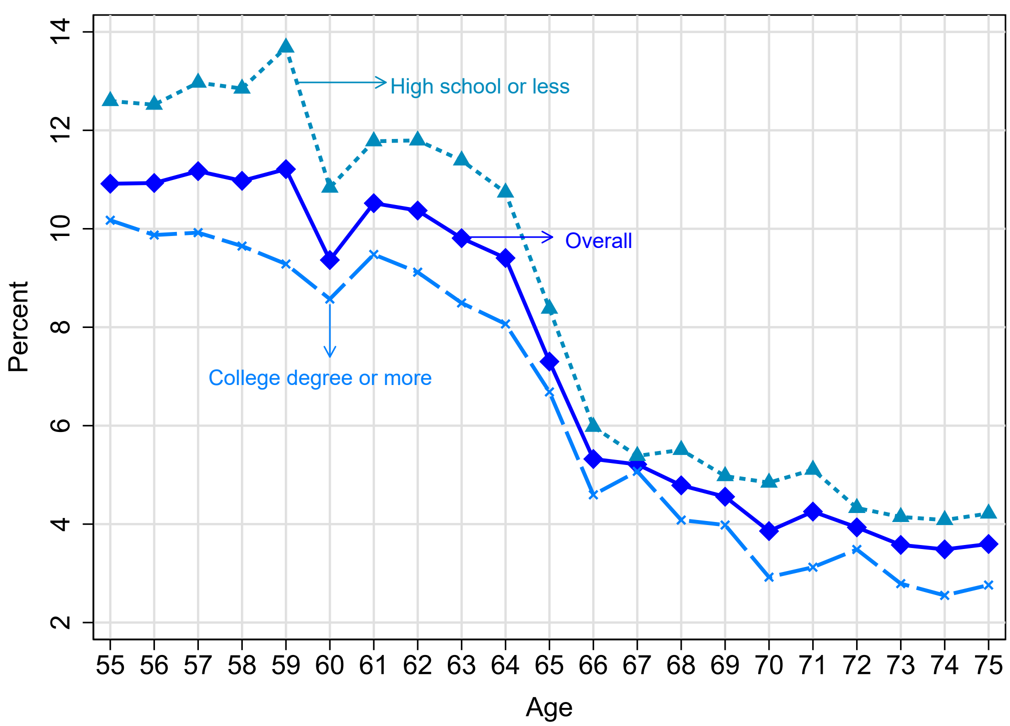


Note: This figure displays the percentage of individual who self-report having health insurance coverage as a function of age and educational attainment. Calculated from NHIS data from 2002 to 2012.

A remarkable feature of the US healthcare system is the discrete and sudden increase in health insurance coverage that occurs at age 65. After age 65, the government provides nearly universal healthcare through Medicare while there is no universal form of coverage for those younger than 65. Figure 1 displays the rate of health insurance coverage as a function of age and education (used as a proxy for income that is not subject to selection concerns around the age 65 threshold; i.e. wealthy individuals may retire or reduce working hours early and report low incomes near age 65) calculated from the National Health Interview Survey. Before age 65, there is a substantial gap in coverage between educational groups of roughly 10 percentage points; however, at age 65 there is a jump in coverage for both education groups and a large convergence in coverage rates due to the sudden availability of Medicare.

The increase in insurance coverage is both quantitatively large and extremely salient — it is a well-known fact among US individuals that Medicare eligibility begins at age 65. As a result, we might expect to see large changes in behavior around the age 65 threshold, particularly for individuals who have no health insurance or who are on cheaper high-deductible health plans and anticipate experiencing large declines in the marginal cost of receiving healthcare upon turning 65.

Figure 2: Delayed Medical Care by Age and Education



Note: This figure displays the percentage of individual who self-report having delayed medical care in the last year for cost-related reasons. Calculated from NHIS data from 2002 to 2012.

Figure 2 displays the percent of individuals in the NHIS who reported delaying healthcare in the last year for cost-related reasons as a function of both age and education level. Unsurprisingly, highly educated individuals report delaying healthcare less often. The percentage of individuals who report delaying healthcare drops substantially from ages 64 to 66 as individuals become eligible for Medicare. Similar to Figure 1, the gap between education levels also shrinks substantially at age 65, consistent with the idea that the increase in insurance coverage (which is larger for the low-educated group) is driving the decline.

One concern with interpreting these results is that changes in Medicare coverage at age 65 may be confounded with a jump in retirement. Figure A.2 displays both health insurance coverage rates and employment rates as a function of age. While health insurance coverage increases substantially at age 65, the employment rate declines smoothly with no sudden changes, assuaging any concerns that changes in working habits or leisure time may be driving the results.

Although it is difficult to make strong conclusions based on responses to survey questions about whether or not individuals delayed care, I interpret these figures as suggestive evidence that a fair number of individuals delay healthcare and that public health insurance can reduce the extent to which individuals delay. This motivates a model of endogenous health expenditure and credit-constrained individuals who face substantial incentives to delay care.

3. Model

The goal of the model is to allow the evaluation of the trade-off between the increase in cost-effectiveness when reducing delayed care and the increase in total outlays that occur due to the subsequent decline in mortality. Each period, individuals face a trade-off between consumption and investment into health. Insurance coverage, either purchased or provided by the government, reduces the marginal cost of health investment and of health emergencies. The expansion of public insurance must be funded by increases in the income tax rate which distorts individuals' labor supply decisions and reduces output.

Because health and mortality occur at the individual level, I model the problem of individuals rather than of households. Although it lacks interesting behavior such as intra-household risk sharing, abstracting from household structure keeps the model tractable and creates a clear link between the model and health data, which are measured for individuals.

Time is discrete and runs infinitely. Individuals are heterogeneous in their income y, savings b, health h, and age a. Exogenous measure n individuals are born at age 18 each period and age increases by 1 each period thereafter. At the end of a period, an individual faces an age- and health- dependent probability of dying $\pi(h, a)$ (discussed in detail later).

3.1. Preferences

Individuals have preferences over streams of consumption $\{c_a\}_{a=18}^{100}$, labor supply $\{l_a\}_{a=18}^{100}$, and mortality risk $\{\pi_a\}_{a=18}^{100}$. I assume that individuals are born at age 18 and die with certainty at age 100 (i.e. $\pi_{100} = 1$). I follow the sparse literature on modeling preferences over mortality by having period felicity be equal to felicity from consumption and labor $u(c,l)$ plus a "joy-of-life" parameter $\bar{u}$ which represents the additional utility an individual receives simply for being alive for the period so that period felicity is given by $\bar{u} + u(c,l)$.

Individuals discount the future exponentially. Death, if it occurs, occurs at the end of each period so that π_a denotes the probability that an individual dies at the end of period in which they are age a and does not live to see age $a+1$. From the perspective of today, an individual of age a sees their future felicity at age $a+1$ as $\beta((1-\pi_a)(\bar{u} + u(c_a,l_a)) + \pi_a \cdot 0)$ (the utility of death is normalized to zero). All together, an individual's present-discounted lifetime utility is given by

$$U(\boldsymbol{c},\boldsymbol{l},\boldsymbol{\pi}) = \sum_{a=18}^{100}[\prod_{j=17}^{a-1}(1-\pi_j)]\beta^a[\bar{u} + u(c_a,l_a)] \tag{1}$$

In essence, period mortality risk π_a acts as a time-varying discount factor by inducing individuals to put less weight on utility from periods that they are less likely to live to see.

3.2. Health and Healthcare Expenditure

An individual's health status is written as a health index h with higher values of h representing healthier individuals. In this way, h can be thought of as a sort of "health capital". When brought to the data, h will correspond to a measured health index that lies in $[0,1]$ with $h=1$ and $h=0$ representing maximally and minimally healthy individuals respectively.

Individuals can increase their health status using medical care, labeled i for health investment, through the health accumulation equation

$$h_{t+1} = (1-\delta_a-\delta_x)h_t + \phi_a i_t^{\psi} \tag{2}$$

where δ_a is the natural rate of health depreciation which may depend on age, δ_x is the (possibly zero) depreciation from an acute emergency shock (discussed in detail in two paragraphs), ϕ_a is an age-dependent productivity parameter governing how effectively dollars of healthcare spending translate into units of health, and ψ is a returns-to-scale parameter that governs how quickly health can be accumulated. The age dependence of δ_a and ϕ_a reflects the fact that the process of health accumulation and maintenance becomes more difficult with age.[1]

This (non-emergency) medical care expenditure represents all spending that is non-urgent and is done largely for the purpose of curing (or potentially preventing) disease. Spending on prescription drugs, such as statins to reduce cholesterol, and spending on (non-emergency) bypass surgery to reduce arterial blockage are both examples of medical spending. This spending is non-urgent in the sense that there are no immediate consequences for an individual who chooses to forgo the spending; however, doing so may cause the individual's health to worsen.

The return to scale parameter $\psi < 1$ acts as an adjustment cost of jumping from unhealthy to healthy by spending heavily in a single period. Instead, health is acquired most efficiently through small investments made each year. Such a notion is intuitive and consistent with data; Hosseini, Kopecky & Zhao (2021) estimate an AR(1) persistence parameter of 0.99 for their measure of health status indicating that health is highly auto-correlated.[2] From the perspective of delayed care, ψ governs that intemporal elasticity of substitution for healthcare spending — that is, how good of a substitution is spending tomorrow for spending today.

Heath Emergencies: Each period, an individual faces an age- and health- dependent risk of experiencing a health emergency, denoted $\pi_x(h,a)$. An emergency carries two consequences. First, it results in an additional one-time depreciation of an individual's health stock, described by δ^x in the health accumulation equation (2). Second, the individual must pay the costs of their emergency healthcare denoted x. These costs are stochastic and, conditional on experiencing an emergency, follow a log-normal distribution with a mean and variance that depend on the individual's age and health.

$$x(h,a) \sim \begin{cases} 0 & \text{with prob. } 1-\pi_x(h,a) \\ \log N\big(\mu(h,a), \sigma(h,a)\big) & \text{with prob. } \pi_x(h,a) \end{cases} \tag{3}$$

Consistent with the literature (e.g. De Nardi, French & Jones 2016), the mean and variance of the shock depend are allowed to depend on an individual's health and age.[3]

[1] One may also wish to consider non-monetary investments into health (e.g. exercise). In Appendix Section B.1, I extend the model to allow for such inputs and show that this extended model is isomorphic to my baseline model under intuitive conditions.

[2] It is worth noting that their persistence estimate is statistically different from 1 suggesting that there is no unit root in health

[3] In Appendix Section B.4, I show evidence of these relationships, although the depends of the variance

While health investment i includes all spending that could be deferred without immediate consequence, emergency expenditure represents urgent spending that cannot be delayed. A straightforward example would be angioplasty administered at the ER to stop a heart attack or emergency surgery for the victim of a severe car crash. Although not technically compulsory, most patients aside from the few who leave the ER, ICU, or otherwise act Against Medical Advice (AMA) treat this spending as effectively compulsory; a doctor prescribes care, and the patient receives the treatment and later pays for it (or discharges the medical debt through bankruptcy).

With the health processes specified, it is apparent there are three reasons that individuals choose to invest in health. First, higher health reduces the likelihood of mortality which represents a direct increase in utility. Second, it reduces the probability of a health emergency which is costly for the individual. Finally, even conditional on having a health emergency, a healthy individual pays lower costs.

Imperfect Information: I allow for a subset of individuals to be poorly informed about the benefits and importance of health and health investment; this ends up being necessary for the model to match aggregate levels of health investment. In particular, I assume that a poorly informed individual with health status h perceives themself as facing the mortality and emergency risks of an individual with health status $h^* > h$. When I bring the model to the data, I measure h using a health index falling in the interval $[0, 1]$ so a natural choice for h^* is

$$h^* = (1 - \chi)h + \chi \tag{4}$$

so that $\chi = 0$ reflects a perfectly informed individual and $\chi = 1$ reflects the extreme that an individual is complete ignorant to the benefits of health.

An individual's information status follows a binary Markov process I and switches between "well-informed" and "poorly-informed" stochastically. Agents are naive about their present and future misinformation (abstracting from the complexities that arise in dynamic imperfect information problems) and never suspect that they might be wrong in the current period or that they might be wrong in the future, even if they are correct today.

3.3. Health Insurance

Individuals can purchase health insurance to help pay for medical expenses and reduce the riskiness of emergency expenditure. Each period t, individuals choose to purchase (or not purchase) exactly one insurance plan from the set of plans for which they

of the shock on an individual's health is statistically insignificant.

qualify. The individual is then covered by that plan in the next period $t + 1$. In this way, individuals may re-optimize their choice of plan each period but must commit to buying a plan before they realize their exact draw of stochastic shocks for the period of coverage.

Every insurance plan p is indexed by a tuple (λ, ν, d, P) representing the plan's copay rate λ, coinsurance rate ν, deductible d, and premium P. Emergency expenditure is covered through a standard deductible-coinsurance system; an individual facing emergency costs m in a single period must first pay up to their deductible d before any insurance coverage kicks in. Then the individual's insurance pays proportion $(1 - \nu)$ of any costs beyond the deductible within the period leaving the individual responsible for paying the remaining fraction ν.[4] In total, the individual's share of emergency costs m is given by $\min(d, m) + \nu \max(m - d, 0)$.

The insurance plan also subsidizes non-emergent care through the copay rate λ. Operating similarly to the coinsurance rate, an individual must pay proportion λ of their medical expenditure i while insurance pays the remaining $(1 - \lambda)$. In addition to being a realistic feature of the model, such a subsidy makes sense from the perspective of the insurance company; the coverage of emergency expenditures introduces a moral hazard problem as individuals no longer face the full cost of their emergency expenditures and no longer fully internalize the benefits of investing in their health, and reducing their marginal cost mitigates this distortion.

Finally, the premium P is the flat per-period cost of the individual's insurance plan. Although the insurance company would like to charge different individuals different premiums, they are restricted top charging a single premium for all consumers (in the context of the US, this may be thought of as occurring through employer-provided insurance charging a single price or marketplace insurance being restricted by community rating rules). Altogether, an individual with insurance plan p, indexed by $(\lambda_p, \nu_p, d_p, P_p)$, who spends i on medical care and faces emergency expenditure m must pay out-of-pocket costs given by

$$\chi_p(i, m) = \underbrace{\lambda_p i}_{\text{Investment}} + \underbrace{\min(d, m) + \nu \max(m - d, 0)}_{\text{Emergency}} + \underbrace{P_p}_{\text{Premium}} \tag{5}$$

Available Plans: There are four available health insurance plans and an option to be uninsured (which is treated symmetrically). The copay rate, coinsurance rate, and deductible (λ, ν, d) for each plan are taken to be exogenous and identical across individ-

[4]For simplicity, I abstract from out-of-pocket maximums, although these could be incorporated in the insurance scheme without much technical difficulty.

uals while the premium P is endogenous (for market-provided plans).

All individuals are eligible to purchase an individual marketplace plan each period, denoted by $p = \text{IND}$. In addition, some individuals are eligible to purchase employer-provided insurance $p = \text{EMP}$. Although there is no *a priori* reason to prefer employer-provided insurance over marketplace insurance, when turning to the data, it is clear that employer-provided insurance plans offer lower deductibles and coinsurance/copay rates, on average, than marketplace plans. Additionally, government subsidies will ensure that, despite better coverage, employer-based plans charge lower premiums than marketplace plans.

Access to employer-provided insurance is not universal however. In reality, only individuals working for an employer who chooses to provide insurance or who previously worked for such an employer and remain covered through COBRA requirements have access to employer-provided programs. Replicating such a process in the model is difficult as the model lacks well-defined notions of job-switching or unemployment and tracking COBRA eligibility would involve many new state variables. Instead, I opt to model eligibility as a simple binary Markov process M.

In addition to the employer-provided and individual marketplace plans, the government administers Medicare ($p = \text{MCR}$) and Medicaid ($p = \text{MCD}$). Medicare is available to all individuals age 65 or older. Although in reality there are many different coverage and plan decisions an individual must make *within* Medicare, I abstract from these and model Medicare as a single insurance plan. Medicaid, the US government public insurance for poor individuals, is made available to all individuals below a productivity threshold. Like with Medicare, I condense the complex reality of multiple Medicaid plans into a single representative plan. As part of this simplification, I model Medicaid availability as a function of productivity rather than earnings, simplifying away any labor market distortions.

Finally, individuals have the option to forgo insurance and enter the next period uninsured which is modeled as a fifth insurance plan with a deductible, copay, and coinsurance all equal to zero. The premium for the uninsurance "plan" is also set by the government and may not be zero, reflecting the presence of an individual mandate which charges individuals for failing to purchase insurance.

Insurance Firms and Pricing: Market provide insurance plans (p$\in$ EMP, IND) are each provided by their own representative insurance firm. These firms take the plan parameters as exogenous and face exogenous load parameters $\kappa_{\text{EMP}}, \kappa_{\text{IND}} > 1$ that summarize the overhead costs of administration a firm that pays out x in total coverage must collect κx in premiums in order to break even for the period. Firms then set prices subject to a zero profit condition (i.e. perfect competition)[5] which is given by

$$P_{\text{EMP}} \int_{p=\text{EMP}} G(a) d\Omega = (1 - s_{\text{EMP}}) \qquad \kappa_{\text{EMP}} \int_{p=\text{EMP}} \int \big(i^* + x - \chi_{\text{EMP}}(i^*, x)\big) df d\Omega \tag{6}$$

$$P_{\text{IND}} \int_{p=\text{IND}} G(a) d\Omega = \qquad \kappa_{\text{IND}} \int_{p=\text{IND}} \int \big(i^* + x - \chi_{\text{IND}}(i^*, x)\big) df d\Omega \tag{7}$$

where Ω is the distribution of individuals across states and f is the (state-dependent) distribution of emergency expenditure shocks x. The LHS of each equation is the firm revenue given by the baseline premium chosen by the firm P_{EMP} and P_{IND} adjusted by the age-specific premium schedule $G(a)$ for individuals who purchases the plan last period. The RHS is the total outlays of the firm multiplied by the loading factor. Finally, s_{EMP} is a proportional government subsidy for employer-based insurance.

3.4. Income and Labor Supply

Individuals younger than 65 years participate in the labor market and supply labor to the healthcare and consumption sectors, earning wage w_h and w_c per efficiency unit of labor supplied to each sector respectively. Individuals have a single measure of labor productivity z which summarizes their total available efficiency units of labor. Labor productivity is given by the following stochastic process (with time subscripts suppressed where possible):

$$\begin{aligned} z(z^p, z^s, a) =& e^{g(a)+z^p+z^s} \\ z^p_{t+1} =& z^p_t \\ z^s_{t+1} =& \rho z^s_{t+1} + \varepsilon_t \\ \varepsilon_t \sim& N(0, \sigma) \end{aligned} \tag{8}$$

Here z^p is the individual's permanent productivity component which is invariant over their life-cycle while z^s represents a stochastic AR(1) component that leads to short-term fluctuations in income. Finally, $g(a)$ is a life-cycle component that depends on age a, allowing for deterministic life-cycle trends in productivity.

An individual who supplies labor l_m to the medical sector and labor l_c to the consumption goods sector earns pretax income given by

$$y_{\text{pre-tax}} = (w_m l_m + w_c l_c) z(z^p, z^s, a) \tag{9}$$

[5]The zero-profit assumption is of little consequence for this particular model. An alternative approach could use a break-even loading factor $\hat{\kappa}$ and a constant markup σ. Because I observe loading factors directly from individual expenditure data, these two parameters are not separately identified, and replacing κ with $\hat{k}(1+\sigma)$ when $\kappa = \hat{\kappa}(1+\sigma)$ changes nothing.

Individuals face disutility from their aggregate labor supply l which is given by a CES-style aggregator of labor supply to the healthcare and consumption sectors

$$l = \left((1-\alpha_m)l_m^{\frac{\xi+1}{\xi}} + \alpha_m l_c^{\frac{\xi+1}{\xi}}\right)^{\frac{\xi}{\xi+1}} \tag{10}$$

While abstract, this reduced-form description of labor supply captures an upwards-sloping relative supply curve for healthcare labor in a tractable way by allowing the relative labor supply decision to be solved analytically, yielding a relative labor supply towards the health sector $\frac{l_m}{l_c}$ that exhibits a constant elasticity of ξ with respect to the relative wage (Appendix Section B.2 provides details).

Taxes and Retirement: Working individuals pay progressive income taxes that are used to fund government-provided health insurance (Medicare and Medicaid) as well as social security payments. After-tax period earnings for an individual younger than 65 are given by

$$y_{a<65}(z^p, z^s, a, l_h, l_c) = T\big((w_h l_h + w_c l_c)z(z^p, z^s, a)\big)$$

where T is a monotonic (but unspecific for now) function. It is important to note that as long as T is monotonically increasing, the solution to the labor allocation problem above does not change.

At age 65, individuals retire exogenously and fix their labor supply l to 0 for the remaining periods of their life. Retired individuals receive social security income which depends on their permanent productivity and is given by the exogenous function $y_{a\geq 65}(z^p)$.[6]

3.5. Consumption, Savings, and the Budget Constraint

Individuals split their income between consumption c (the numeraire), medical expenditure i, emergency expenditure x, and assets b, which are risk-free and pay an interest rate of r_t each period. Markets are incomplete, and individuals cannot borrow (assets b_t must be weakly positive). The budget constraint of an individual with assets

[6] In reality, US social security income is determined by one's entire earnings history, but faithfully modeling this requires an intractible number of additional state variables. Following much of the literature, permanent income provides a good approximation of this average for everyone except the ex-post luckiest and unluckiest individuals who earned substantially more or less than their permanent income. Given the large increase in tractability for relatively little loss in accuracy, I opt for a simple model of retirement income.

b_t, insurance plan p_t, and productivity z is given by

$$c_t + b_{t+1} + p_h \chi_p(i_t, m_t) = (1 + r_t) b_t + T\big((w_{h,t} l_{h,t} + w_{c,t} l_{c,t}) z\big) \qquad \text{if } a < 65 \tag{11}$$
$$c_t + b_{t+1} + p_h \chi_{\text{MCR}}(i_t, m_t) = (1 + r_t) b_t + y_{a \geq 65}(z_t^p) \qquad \text{if } a \geq 65$$

where p_h is the relative price of healthcare goods.

3.6. The Individual Optimization Problem

I relegate the full specification of the individual optimization problem and definition of equilibrium to Appendix Section B.3. Here, it is sufficient to point out that there are eight individual-level state variables. They are assets b (1), health h (2), age a (3), permanent and temporary productivity (z^p, z^s) (4-5), and finally insurance plan, employer-provided insurance eligibility, and information status (p, e, χ) (6-8). Of course individuals also face the cross-sectional distribution of individuals across states Ω as an aggregate state variable. One major advantage of this modeling approach to health is that framework above maps cleanly into the two asset framework popularized in the Heterogeneous Agent New Keynesian (HANK) literature and thus can leverage the computational advances aimed towards these models.

3.7. Delayed Medical Care

How does delayed care arise in the model? Like all consumption-savings models, individuals use their assets b to smooth their consumption. Within-period optimization between consumption and health spending dictates equality between the marginal benefit of each. The first-order condition is

$$u_c(c^*, l^*) = \frac{\phi_a \psi i^{\psi - 1}}{p_h \lambda_p} \beta \big(1 - \pi(h, a)\big) V_{h'}(h'^*) \tag{12}$$

where I have suppressed most of the inputs into the value function for brevity.

From equation 12, it is clear that as long as V exhibits diminishing marginal returns to health (as is the case in the estimated model), the marginal return to additional health $V_{h'}(h'^*)$ is inversely related to the individual's copay rate λ_p (as the individual acts to equate marginal benefits with marginal costs). Additionally, the envelope condition for the marginal value of health

$$V_h(h) = \underbrace{-\beta\pi_h(h,a)\mathbb{E}\big(V(h')\big)}_{\text{Reduction in mortality}} + \underbrace{\beta\big(1-\pi(h,a)\big)\frac{\partial}{\partial h}\mathbb{E}\big(V(h')\big)}_{\text{Reduction in Emg. Risk}} + \\ + \underbrace{(1-\delta_a-\delta_e)\beta\big(1-\pi(a,h)\big)\mathbb{E}\big(V_{h'}(h')\big)}_{\text{"Health Tomorrow"}} \quad (13)$$

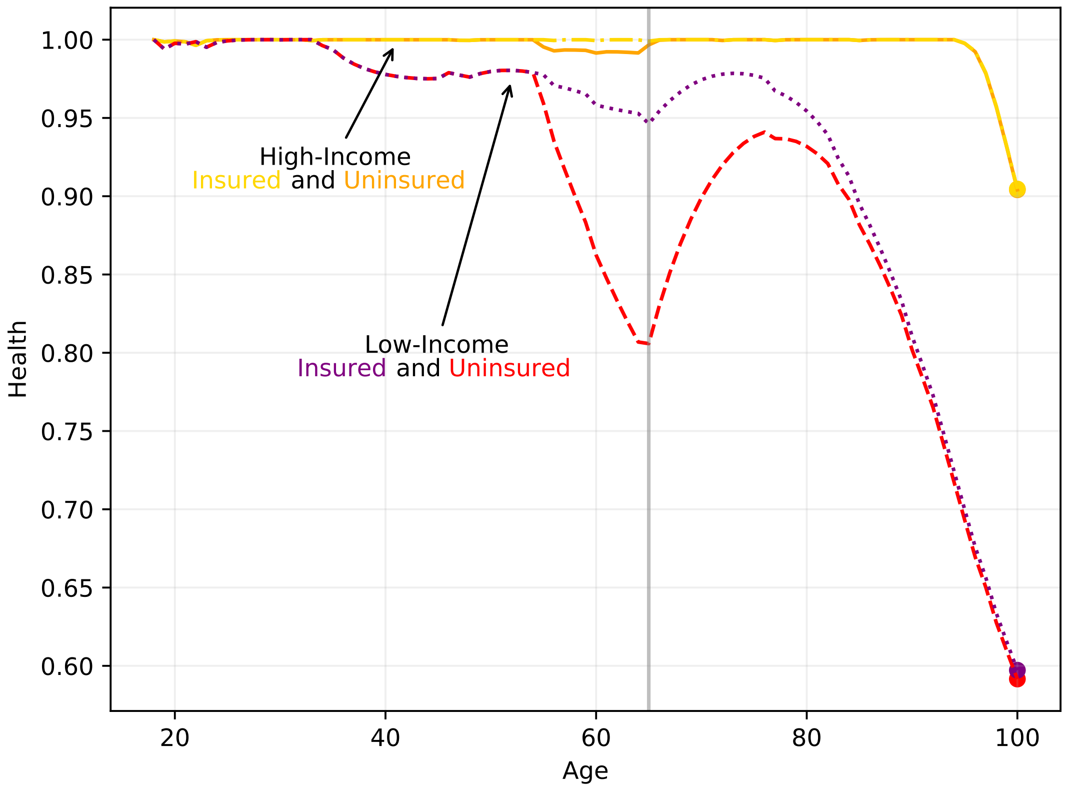
reveals that health is a forward-looking asset; part of the benefit of being healthy today is that one will continue to be healthy tomorrow. Iterating this relationship forward through time, it is clear that the value of health today depends on the value of health t periods in the future discounted by $(1-\delta_a-\delta_e)^t\beta^t\prod_{i=0}^{t-1}(1-\pi(a+i,h_i))$.

Delayed care arises from the interaction of these two relationships. The individual expects to receive health insurance in the future, lowering λ_p, and thus the future marginal value of health. Because the value of health today depends on the value of health in the future, especially as the individual approaches the period they will receive insurance, this lowers the value of health today, and the individual reduces spending. The economic intuition is straightforward. If health is going to be so much cheaper tomorrow (or in two or three periods) and a large portion of the value of health comes from its continuation value, why bother investing in health today? The most effective strategy is to treat health as an asset to be run down and then replenished once covered by insurance. This incentive is mitigated by the decreasing returns to medical care each period which ensures that medical care tomorrow is not too good of a substitute for medical care today, but it is still strong enough to generate quantitatively important behavior.

A Quantitative Example: Figure 3 provides a quantitative example of individual health investment behavior. The figure displays individual health (on the y-axis) over the lifecycle of the individual as measured by their age (on the x-axis) for four different individuals. These individuals all receive an identical series of income and health shocks (with mortality shocks set to zero to prevent different years of death from precluding comparisons) but differ by their permanent income and their access to employer-provided health insurance.

The dynamics of health investment can be seen clearly in the figure. Early in life, the four individuals receive the same shocks but, because they spend more on healthcare, the high-income individuals recover from shocks almost immediately and maintain high average health. Insurance status makes little difference in health dynamics at this point. Later in life, around age 58, the four individuals experience a series of bad

Figure 3: Life-cycle Health Dynamics for Four Individuals



Note: This figure displays life-cycle health dynamics for four individuals who receive an identical series of health and income shocks. The yellow line displays the health of an individual with high permanent income and access to employer-based insurance. The orange line displays the same for an individual with high permanent income and no health insurance. The purple and red lines display the health of low permanent income individuals with and without access to employer based insurance respectively.

health shocks just before retirement. Here the dynamics begin to diverge. As was the case with shocks early in life, the high-income individuals spend more on healthcare, recover from shocks faster, and keep their health at a higher level than the low-income individuals. However, as the individuals are approaching the age 65 threshold after which they will receive Medicare, they begin to delay care.

This delay of care is most noticeable in the differences between insured and uninsured low-income individuals. The insured individual (purple) continues to invest in health and mitigates the effects of the shocks to some extent. The situation is more dire for the uninsured individual (red) who spends almost nothing on healthcare and experiences substantial declines in health. After these individuals turn 65 and receive coverage through Medicare, their health slowly converges and the gap disappears around age 85 due to the uninsured individual's higher spending. Appendix Figure A.1 shows this directly by displaying annual healthcare spending for each of these two individuals over their life-cycle.

3.8. Production and Government Budget

With the household problem finished, closing the economy requires specify firm and government behavior, both of which are simple. There are two representative firms producing healthcare (m) and consumption (c) goods respectively, setting output and input prices according to perfect competition, and operating the technologies

$$
\begin{aligned}
p_m Y_m =& p_m A_m L_m^{\alpha} K_m^{1-\alpha} \\
Y_c =& A_c L_c^{\alpha} K_c^{1-\alpha}
\end{aligned}
\tag{14}
$$

where p_m is the price of healthcare good. Capital can flow freely between sectors so that aggregate demand for capital is simply given by $K = K_m + K_c$. As a result, the rate of return on capital is equalized between the two sectors $r_m = r_c$, justifying the modeling decision that households have only a single asset in which to invest.

The government spends on social security, Medicaid (for the poor), Medicare (for the old), and the subsidy for employer-provided insurance. The government also spends an exogenously-specified amount on a generic public good G that provides no utility. Taxes are collected through the income tax function T; following Heathcote, Storesletten & Violante (2017), T takes a simple monotonic form $\lambda_\tau y^{1-\tau}$. The government cannot borrow and instead adjusts the level parameter of the tax function λ_τ to ensure that the budget balances each period.

4. Data and Model Estimation

Bringing quantitative discipline to this model involves estimating many different parameters; rather than provide an exhaustive description of how each parameter is estimated, this section summarizes the data and quasi-experimental results used for estimation, discusses the key parameters driving the main quantitative results, and provides intuition for how the data used in estimation discipline these parameters. A thorough discussion providing estimation details for every parameter can be found in Appendix Section B.4.

4.1. The Medical Expenditure Panel Survey

The Medical Expenditure Panel Survey (MEPS) serves as the primary source of data for quantification of the model. These nationally-representative data contain detailed information on individual healthcare expenditure, insurance coverage and plan details, and health status. The panel structure of the MEPS allows observation of individuals' eventual health outcomes (at a time horizon of up to two years, after which they drop out of the rolling panel) conditional on their characteristics such as health or age. This allows direct estimation of many important functions in the model, such as the annual mortality probability or the probability of having a health emergency as a function of health and age.

A second useful feature of the MEPS is that the data on healthcare expenditures are largely measured using billing data from healthcare providers, rather than estimates from surveyed individuals.[7] These data reflect actual payments made to the provider and thus are not subject to rounding bias (as is common in survey data) or issues of nominally inflated medical bills that are later reimbursed by insurance companies at lower rates. Additionally, these data distinguish between sources of payment, allowing the separation of out-of-pocket costs and costs covered by insurance. This allows direct estimation of coinsurance rates and copayment rates for representative insurance plans.

Bringing expenditure data to the model requires making a distinction between spending on health investment (i) and spending on emergency health events (x). This is difficult to do even conceptually; an emergency bypass surgery also cleans arteries and results in a long-term improvement in health. I opt to categorize any spending that occurs in an emergency room or inpatient hospital setting as emergency spending and any other spending as investment, largely because this is unambiguously observed in the data. Appendix Table A.2 reports basic summary statistics for both investment and emergency spending. The overall patterns of spending are not surprising. Health emergencies are relatively rare with most individuals reporting no emergency expenditure in a given year regardless of age. Both types of spending are larger for older individuals and exhibit a strong right skew with means larger than medians.[8]

Measuring Individual Health: The MEPS also provides information on individual health status and outcomes. Mapping the complex reality of health to a simplified, abstract concept amenable to economic modeling is a classic difficulty in the macro-health literature. One solution has been to restrict health to a small number of discrete cate-

[7]Unfortunately, not all spending from all individuals is taken from billing data. Whether billing data is collected for a health event depends on the insurance coverage of the individual and the provider seen. Appendix Table A.1 summarizes the fraction of spending taken from billing data for two key categories of provider: office-based physicians and hospitals. The MEPS attempts to use information from the surveyed providers to impute individual-reported spending at providers who are not surveyed. The details on this process are sparse and the public-use data lack imputation flags, making it impossible to compare imputed spending to that taken directly from billing data. Still, I take the reported spending data at face value. See Cohen (2003) and Zuvekas & Olin (2009) for more details.

[8]These patterns, as well as other results based the separation of spending into investment and emergent, are largely robust to alternative definitions such as limiting emergency expenditure to only that which occur in an emergency room.

gories such as "Good" and "Bad" or ranging from "Excellent" to "Poor" (e.g. De Nardi, French & Jones 2016, Yogo 2016). Particularly in the case of the latter, these categories are often self-reported subjective measures of health that may or may not be related to an individual's actual health (see Spitzer & Weber 2019, for an example).

To overcome these issues, I use the frailty index of Hosseini, Kopecky & Zhao (2021) which has the advantage of being largely objective and close-to-continuous, mapping cleanly to health h in the model. Frailty is constructed from a wide variety of yes or no questions about an individual's health ranging from diagnoses (Have you ever been diagnosed with diabetes?) to cognitive limitations (Do you experience confusion or memory loss?) to common metrics known as Activities of Daily Living (Do you have difficulty getting dressed by yourself?), supplemented with some other objective measures of health such as an indicator for if the individual's BMI is greater than 30 and the individual's K6 score (a common measure of mental health). An individual's frailty is determined by summing the number of "yesses" (referred to as the total number of health deficits) reported by the individual and normalizing by the total number of questions so that the index falls between 0 and 1.

Subtracting the frailty index from 1 creates an analogous health index in which a value of 1 represents a maximally healthy individual who reports having no health deficits and 0 represents a minimally healthy individual who reports having every health deficit. I discuss the distribution of health (displayed in Figure 6) in more detail in a later section on model validation. Here, it suffices to mention that the dsitribution of health exhibits a strong left-skew with a high concentration of healthy individuals possessing health indices between 0.9 and 1.0 with a thin tail on the left-hand side. Very few individuals accumulate an extensive number of health deficiencies.

Hosseini, Kopecky & Zhao (2021) discuss at length the usefulness of frailty as a measure of health and show that it is a strong predictor of many health outcomes, including medical expenditure and mortality, and that it outperforms self-reported measures of health. Appendix Table B.3 replicates these findings and shows that the health index strongly predicts individual mortality, the probability of positive emergency expenditure, and the total amount of emergency expenditure conditional on positive expenditure.[9] The estimate coefficient for health is robust to the inclusion of a wide variety of controls including family income, race, sex, and geographic region, providing suggestive evidence that the index is not picking up variation in non-health-related latent variables.

[9] These regressions also serve as the functions $\pi(h, a)$, $\pi_x(h, a)$, and $\mu(h, a)$ in the model.

4.2. Key Parameters and Model Estimation

An exhaustive list of model parameters and their estimated values can be found in Appendix Section B.4 (Appendix Tables B.1, B.2, and B.4 in particular). Here I discuss the key parameters driving the model results and their quantification.

Preferences: Period utility takes the King-Plosser-Rebelo (KPR) functional form given by

$$u(c,l) = \bar{u} + \frac{1}{1-\sigma} c^{1-\sigma} \left(1 - \kappa(1-\sigma) l^{1+\frac{1}{\nu}}\right)^{\sigma} \tag{15}$$

These preferences exhibit a constant Frisch elasticity of labor supply (chosen to be unity) and, as a result, are commonly used in analyses of general equilibrium responses to income taxation (see e.g. Trabandt & Uhlig 2011). As discussed previously, the parameter $\bar{u}$ governs individuals' willingness to trade off between consumption and mortality and is chosen to match an average Value of Statistical Life (VSL) of 10 million USD (Viscusi & Aldy 2003).

It is important to mention the value of σ, the coefficient of relative risk aversion, which is set at 2. This value is commonly used in both the consumption-savings and taxation literatures. However, Murphy & Topel (2006) and Hall & Jones (2007) demonstrate that this parameter plays a critical role in determining how an agent's willingness to trade-off consumption and mortality changes with income. Intuitively, this occurs because the intertemporal elasticity of substitution governs how much the agent values living to see future consumption. If the IES is high, consumption today is a good substitute for consumption tomorrow, and the agent can consume heavily today and not worry about whether or not they will be alive to see tomorrow's consumption. Conversely, a low IES agent values living to see tomorrow very highly as there is no good substitute for tomorrow's consumption. Thus the CRRA/IES parameter is doing "triple duty"; it governs the agent's risk tolerance (as in the consumption-savings literature), their labor supply response to income shocks (as in the taxation literature), and the income elasticity of their VSL (as in the sparse macro-health literature). The chosen value of 2 gets all three behaviors roughly correct.[10]

Insurance Plans: Condtional on the preference parameters and the distribution of health shocks (which are estimated directly from MEPS data), individuals' willingness to purchase health insurance depends on the parameters of insurance plans — namely the coverage provided and the loading factors dictating premiums. The parameters of representative insurance plans are estimated directly from expenditure data. For exam-

[10] As an alternative approach, one could use a variant of Epstein-Zin preferences, such as those suggested in Córdoba & Ripoll (2017), to adjust parameters governing each behavior separately, but this adds substantial computational complexity for little gain in accuracy.

ple, the copayment rate for employer-provided insurance is estimated by subtracting the percentage of non-emergency spending that is paid out-of-pocket from 1, yielding the percentage of spending that is covered by insurance in practice. Coinsurance rates and rates for marketplace and Medicaid insurance plans are estimated similarly (with the caveat that coinsurance rates must account for deductibles). The parameters for the Medicare insurance plan are set to their statutory values.[11] Loading factors κ_{EMP} and κ_{IND} are estimated similiarly by dividing total premiums collected by total outlays for by employer-provided and marketplace-purchased plans.

Appendix Table A.3 displays the estimated values for all insurance plan parameters. The estimated copay rates are 28 percent and 38 percent for employer-provided and marketplace plans respectively, meaning that covered individuals pay for 28 or 38 percent of their non-emergency medical spending. Coinsurance rates are lower at 10 and 13 percent. The estimate load factor for the marketplace plan is 1.3 compare to 0.7 for the employer-provided plan (employer-provided plans pay out more in benefits than they collect in premiums, at least from households). I interpret the difference in loading factors to be result of government subsidies rather than structural advantages in administration for employer-provided plans. In other words, subsidies cover the remaining 60 percent of payments that would need to be collecting in premiums resulting in an estimated subsidy rate of about 46 percent. These results are in line with typical intuition; employer-provided plans have better coverage and do so with lower premiums (largely a result of subsidies).

The Health Production Function and Quasi-Experiments: The aggregate effects of reducing delayed care through public insurance expansion occur through two primary channels: lower mortality and lower long-run medical costs. The parameters of the health production function (namely the productivity parameter ϕ and the returns to scale parameter ψ) end up being the key parameters governing the quantitative strength of these channels. The value for each parameter is pinned down by a closely related quasi-experimental result, lending credibility to the quantitative predictions of the model.

Intuitively, the health investment productivity parameter ϕ is tightly linked to the mortality channel. A low value corresponds to a world in which spending on healthcare does not lead to very substantial gains in health, and shocks end up being a larger determinant of health than spending. As a consequence, delaying healthcare spending

[11]Fixing parameters to their statutory rates is possible for Medicare but not Medicaid due to the fact that the latter exhibits substantial variation in coverage from state to state. Thus estimating a representative Medicaid plan is dramatically simpler than attempting to somehow aggregate the various state-level plans.

has little impact on mortality, and reductions in delayed care do not lead to substantial reductions in mortality. Conversely, a high value of ϕ corresponds to a world where health is primarily determined by health investment, and delaying healthcare has large effects on mortality.

I leverage this relationship to discipline ϕ by estimating the model to match a quasi-experiment that estimated the impact of public health insurance (Medicaid) provision on mortality. Miller et al. (2021) use the fact that the 2014 Affordable Care Act (ACA) increased the income threshold in Medicaid eligibility only in certain states. Combined with administrative social security data on mortality, they use this variation in a diff-in-diff framework comparing states that expanded to those that did not expand and estimate that Medicaid expansion reduced mortality for newly eligible individuals between 55 and 64 years old by 6.3 percent in the year immediately following expansion and 9.4 percent overall. Including this first moment in model estimation (and leaving the second for validation) disciplines the health production function, as described above, and helps ensure that the model makes accurate predictions about the effects of expanding public insurance.

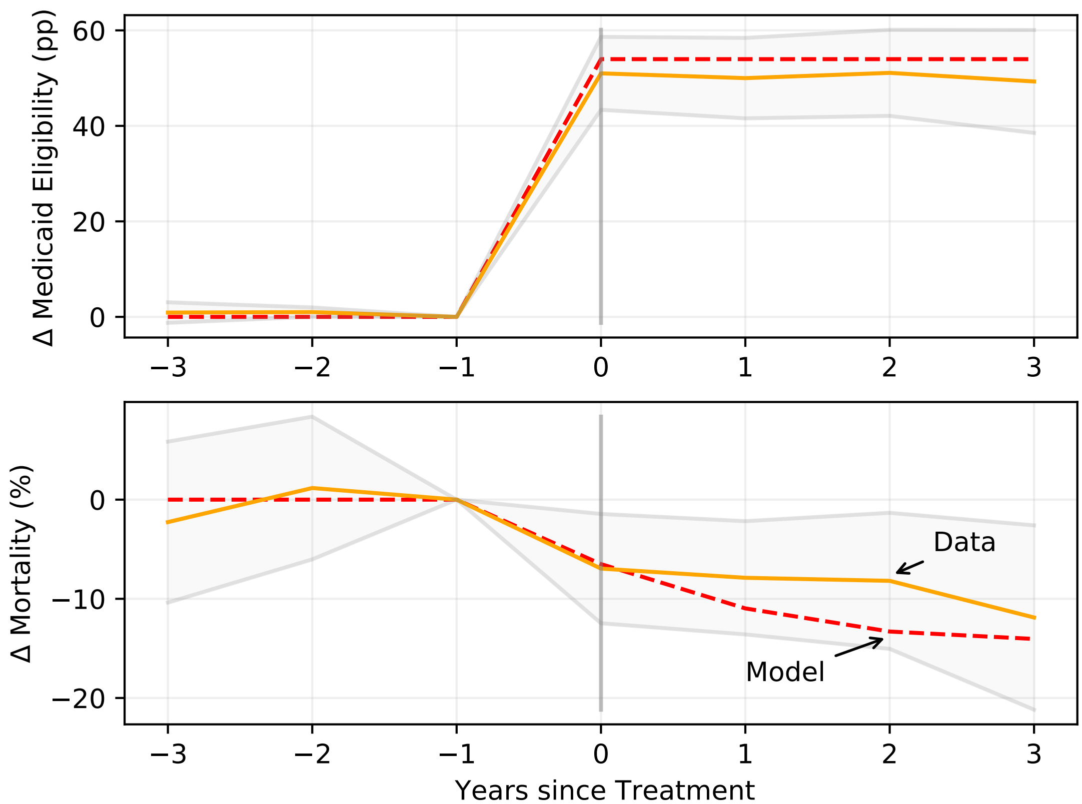
Estimating the model to match this moment requires producing a model analogue for the diff-in-diff estimation strategy. Starting with the steady-state distribution of individuals from a version of the model with a low Medicaid eligibility cutoff (denoted $\underline{z}_{\text{PRE}}$), I replicate the sample selection procedure of Miller et al. (2021) and restrict the sample to include only those who are between 55 and 64 years old and who would qualify for Medicaid under the post-expansion eligibility cutoff $\underline{z}_{\text{POST}}$. I then consider two different worlds, one in which the eligibility cutoff remains at $\underline{z}_{\text{PRE}}$ (control) and the other in which it is increased to $\underline{z}_{\text{POST}}$ (treatment), and simulate the sample forward in each scenario.[12] The model equivalent of the diff-in-diff estimator is constructed by taking the simple difference between mean outcomes in the treatment and control worlds (as the "pre-period" outcomes are identical between treatment and control by construction).

The final detail is choosing appropriate values for $\underline{z}_{\text{PRE}}$ and $\underline{z}_{\text{POST}}$. The latter is straightforward; states that expanded Medicaid under the ACA were required to increase the income eligibility cutoff to 138 percent of the Federal Poverty Level (FPL), making this a natural choice for $\underline{z}_{\text{POST}}$. The choice of pre-expansion cutoff is more complicated as cutoffs for Medicaid eligibility varied dramatically across states before the ACA. In order to maximize comparability between model and data, I opt to set $\underline{z}_{\text{PRE}}$ so that the increase

[12] For the simulation using the expanded eligibility cutoff $\underline{z}_{\text{POST}}$, I assume that the cutoff is increased for all individuals in the economy and, consequently, that the path of prices faced by sample individuals changes. Such an assumption seems sensible in this context where variation occurs at the state level.

in Medicaid eligibility for the treated group in the model matches that estimated in the data (roughly 50 percentage points).

Figure 4: Effect of Medicaid on Mortality: Model vs Data



Note: This figure displays the impact of Medicaid expansion on Medicaid qualification and mortality of low-income 55-64 year old adults. The solid orange line displays the effects estimated in Miller et al. (2021) using a diff-in-diff design. The grey bands represent the estimated 95% CI. The dotted red line displays the change in mortality in the calibrated model.

Figure 4 displays the results of the diff-in-diff estimation in both the data (orange solid line with 95 percent confidence intervals shown in gray) and the model (red dotted line). The top panel displays the increase in Medicaid eligibility which is well-matched by construction. The bottom displays the estimated decline in mortality for the first three periods after expansion (the longest horizon reported in Miller et al. 2021). The model closely matches in initial decline in mortality (unsurprising, as this was a targeted moment). It also performs reasonably well at reproducing the dynamics of the decline in mortality, slightly over predicting the point estimates one and two years after treatment and closely matching the decline after three. This growing decline is intuitive as every year that passes is an additional year that treated individuals will build health relative to control individuals and also highlights one advantage of model-based analysis — the model can be use to analyze the full long-run effect expansion which may be much larger than the impact estimated in the first few years.

While ϕ is tightly linked to the mortality channel, the returns-to-scale parameter ψ governs the cost-savings channel. This parameter acts as the intertemporal elasticity of substitution for health spending. A low value indicates that low health spending today cannot be easily offset by higher health spending tomorrow and that achieving the same level of health through low spending in one period and higher spending in the next requires greater total spending. Consequently, a low value for ψ indicates that reductions in delayed care due to public insurance expansion can potentially lead to large reductions in aggregate healthcare expenditure.

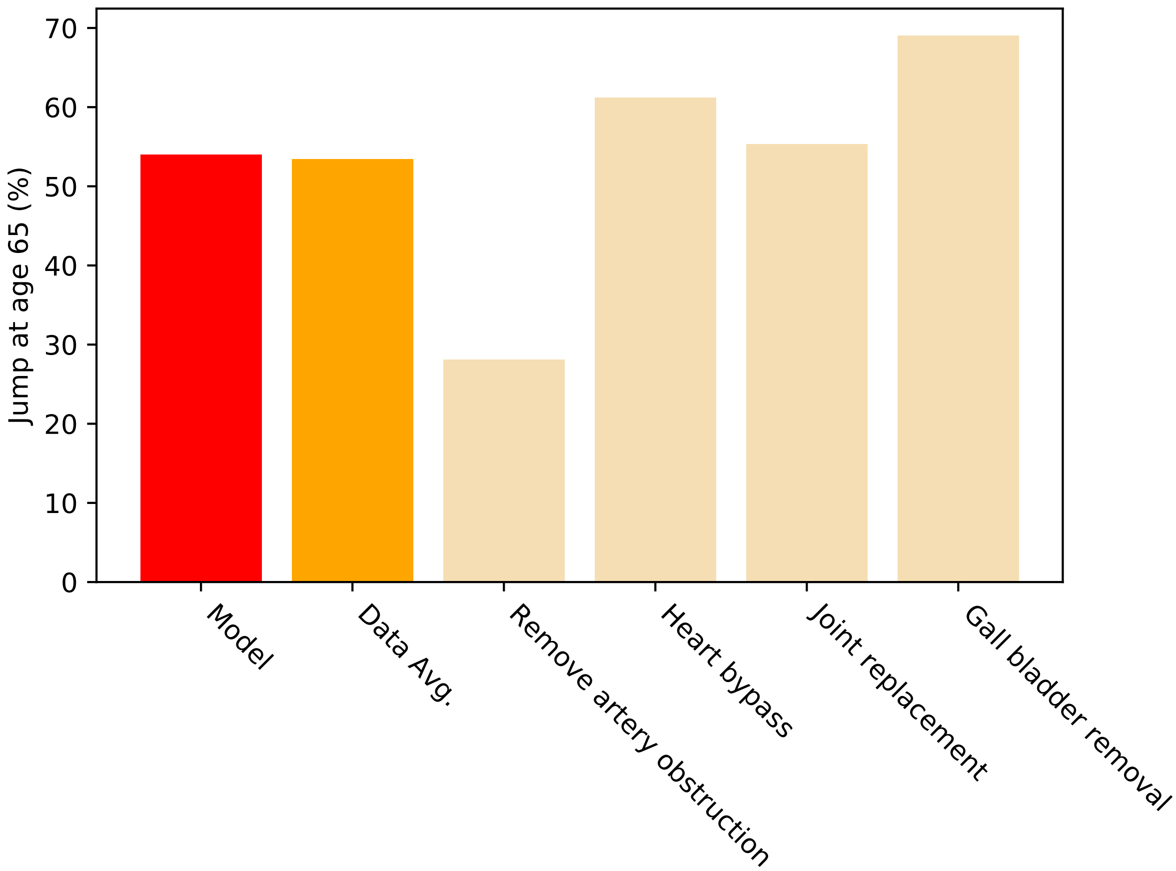
The main source of quantitative discipline for ψ comes from Card et al. (2008) who use a regression discontinuity design to estimate that the use of various healthcare services jumps dramatically at age 65 (when individuals begin to qualify for Medicare).[13] Including this moment in model estimation has the obvious advantage of ensuring that the model matches observed levels of delayed care, but it also turns out it is closely linked to ψ. If the value of ψ is low and delaying care is very costly, then individuals will be less aggressive in their delay of care and the measured jump will be small. Conversely, a high value of ψ means that the consequences of delaying care are small and individual's will choose to high levels of delayed care. This relationship demonstrates an unintuitive "paradox" — the more delayed care that is observed, the lower the (implied) cost of any given quantity of delayed care, making it hard to evaluate its aggregate consequences without a quantitative model.

To perform the RD estimation in the model, I start with the distribution of individuals from the same pre-expansion steady state as the diff-in-diff estimator, as the data used in Card et al. (2008) are from well before the ACA. I then compute health investment spending for each individual in the distribution and estimate a basic RD framework, replicating the authors' methodology as closely as possible.[14] This procedure is complicated by the fact that that authors do not explicitly estimate a jump in expenditure (which would be the ideal comparison to the model) and instead estimate jumps in

[13]While these estimates are more dated than is ideal, they are produced using high-quality, private-use administrative data from hospitals and comparable public-use data do not exist, making it impossible to replicate the methodology on more recent data.

[14]A notable exception is that the authors estimate a quadratic model while I estimate a log-linear model. The estimated values using both approaches are similar in the final estimated model, but the log-linear specification leads to much smoother relationships between ψ and the moment of interest substantially easing estimation.

Figure 5: Jump in Medical Expenditure at age 65: Model vs Data



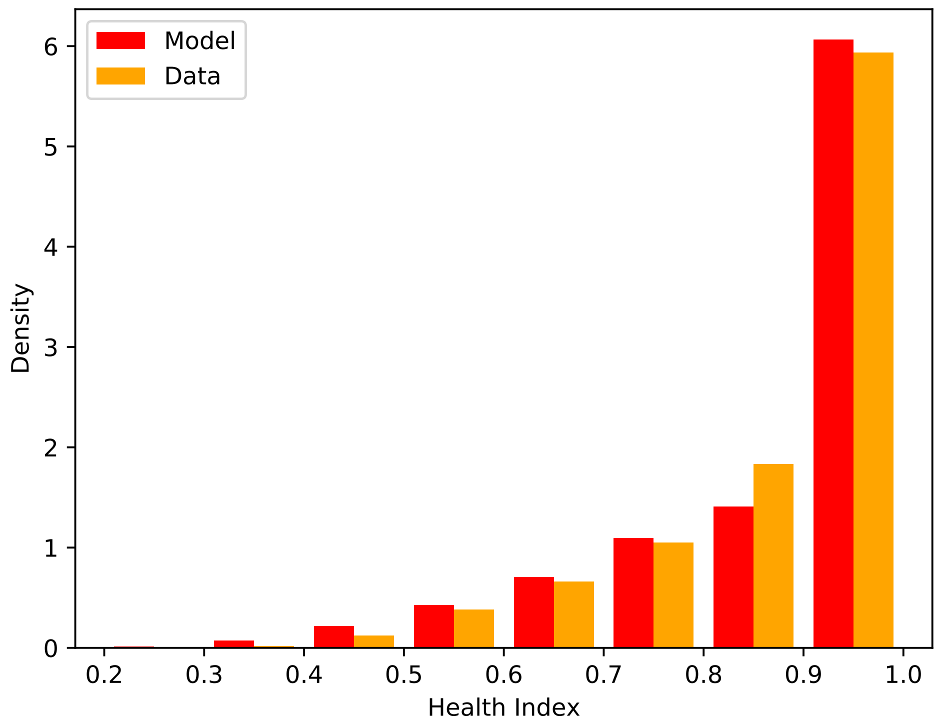
Note: This figure displays the jump in various costly non-emergent medical procedures at age 65 estimated in Card et al. (2008) in tan. The orange bar represents the unweighted average of these four estimates. The red bar represents the jump in average medical expenditure in the pre-Medicaid expansion steady-state of the calibrated model.

the utilization of various procedures separately. To address this, I choose a set of four common procedures and impute the average jump in expenditure as the simple average of the jump in utilization for the four procedures (54 percent). The procedures, their estimated jumps, and the average jump in both the data and the model are displayed in Figure 5.

4.3. Discussion of Parameter Values and Model Validation

As a first pass at model validation, Figure 6 compares the distribution of health in the model to that observed in the MEPS using a pair of overlapping histograms with the red histogram corresponding to health in the model and the orange histogram correspond to the MEPS. Although the mean of this distribution is the only moment targeted in model estimation, it is clear that the model successfully replicates the stark features of the data distribution, including bunching at the top and a thin left tail. The model slightly overpredicts the variance of the distribution as there are more individuals with extremely low health — $h \in (0.3, 0.5)$ — in the model than are observed in data. Given the natural and (relatively) parsimonious description of the health accumulation process, I view this as a substantial success of the model.

Figure 6: The Distribution of Health in the Data and Model



Note: This figure displays the distribution of health in the Medical Expenditure Panel Survey, as measured by the health index based on Hosseini et al. (2021), in orange. The distribution of health in the calibrated model is displayed in red. See the discussion in Section 4.3 for more details.

It is also worth checking if the model produces reasonable values for various aggregate statistics that are relevant to public health insurance coverage and costs. Most importantly, the model is able to deliver accurate predictions for the cost of public health insurance expansion under the 2014 ACA. Although I postpone the details of how this expansion is simulated until the next section, the estimated model implies that such an expansion requires an annual increase of \$60 billion in Medicaid expenditure. In reality, expansion increased expenditure by \$69 billion in 2017 (CMS 2018)[15]; the model

[15] Although the model is largely estimated on data for 2018, actuarial reports from CMS are only publicly available up to 2017 — hence the slight discrepancy in years.

matches the data quite closely

The model also closely matches Medicaid enrollment rates. In the pre-expansion steady-state of the model, 8.2 percent of individuals (i.e. adults) qualifying for and enroll in Medicaid. In 2013 (the year before expansion), CMS (2014) reports that the analogous rate in data is between 6.1 and 10.2 percent .[16] In the post-expansion steady-state, 16.9 percent of individuals are enrolled in Medicare. Although not directly comparable to the data as not all states expanded insurance access, a back-of-the-envelop adjustment based on the fact that roughly two-thirds of the population live in states that expanded access suggests that 13.9 percent ($= \frac{1}{3}8.2 + \frac{2}{3}16.9$) is a reasonable value to compare to data. The observed enrollment rate is between 11.0 and 15.2 percent (CMS 2018); thus the model is again able to match the data.

Finally, we can compare the model's predictions for life expectancy to actuarial predictions made by the US Social Security Administration. The model produces an average life expectancy at age 18 of 84 while the equivalent number from SSA tables is 75 (SSA 2023). Thus the model somewhat over-predicts average lifespans. More important, at least for the quantitative impact of changes in mortality, is the difference in life expectancy between those who qualify for Medicaid and those who do not. In the model, this difference is 5 years. While no direct data analogue exists, Chetty et al. (2016) report that life expectancy in the top 75 percent of the income distribution is 5.5 years higher than in the bottom 25 percent (7 years for men and 4 years for women), indicating that the model is roughly correct on this dimension.

Values of Health Parameters: The productivity and returns to scale parameters of the health production function (ϕ_a and ψ) are both novel and central to the point of the paper, so it is worth briefly discussing them. The estimated values of ϕ_a range linearly from 0.221 at age 20 to 0.097 at age 65 and the returns to scale parameter is estimated to be 0.487. As an intuitive example, these estimated values imply that a 40 year old must spend roughly 3000 USD in order to erase one health deficit.[17] For a 65 year old, the required spending is 9000 USD (combined out-of-pocket and insurance-covered).

Further, the estimated value of ψ implies that these costs scale roughly quadratically; if an individual ignores their first deficit and develops a second, removing both deficits requires either spending four times as much or spreading the expenditure over two

[16]The ambiguity in this range stems from the fact that some adults qualify for Medicaid through disability requirements rather than income requirements. While conceptually the closest analogue to the model would only count adults who qualify through income requirements, I do not observe whether or not a disabled individual would qualify through income requirements were they not disabled. Given the large income penalties associated with disability, it is not hard to imagine that there are many such individuals, making it impossible to pin down the rate more precisely.

[17]One deficit = 0.042 units of h

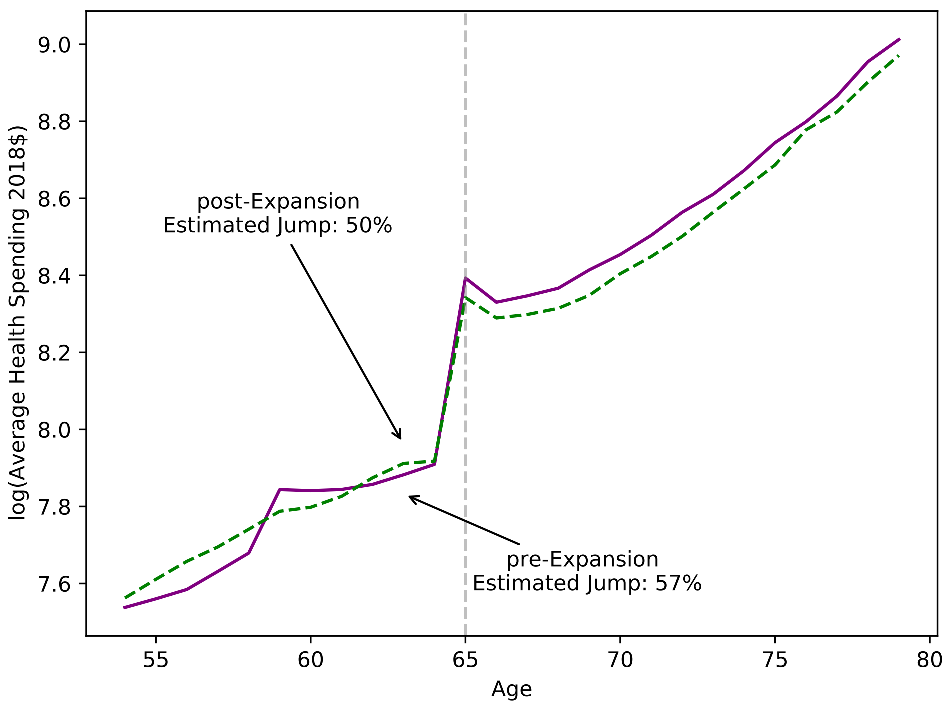
years. While there is no direct data analogue to compare these values against, they seem at least reasonable and roughly in line with intuition.

5. Quantitative Results

I use the estimated model to evaluate the results of an expansion in public health insurance similar to the 2014 ACA Medicaid expansion. Starting from the model steady-state under a Medicaid productivity cutoff of $\bar{z}_{\text{PRE}}$, I increase the cutoff to $\bar{z}_{\text{POST}}$ and simulate the economy forward. The values for $\bar{z}_{\text{PRE}}$ and $\bar{z}_{\text{POST}}$ are taken directly from the model estimation exercise of the previous section. Most of the discussion and results focus on comparisons between steady-states with only a brief mention of transition dynamics.

5.1. Reduction in Delayed Care

Figure 7: Model Implied Delayed Care Pre- and Post- Expansion

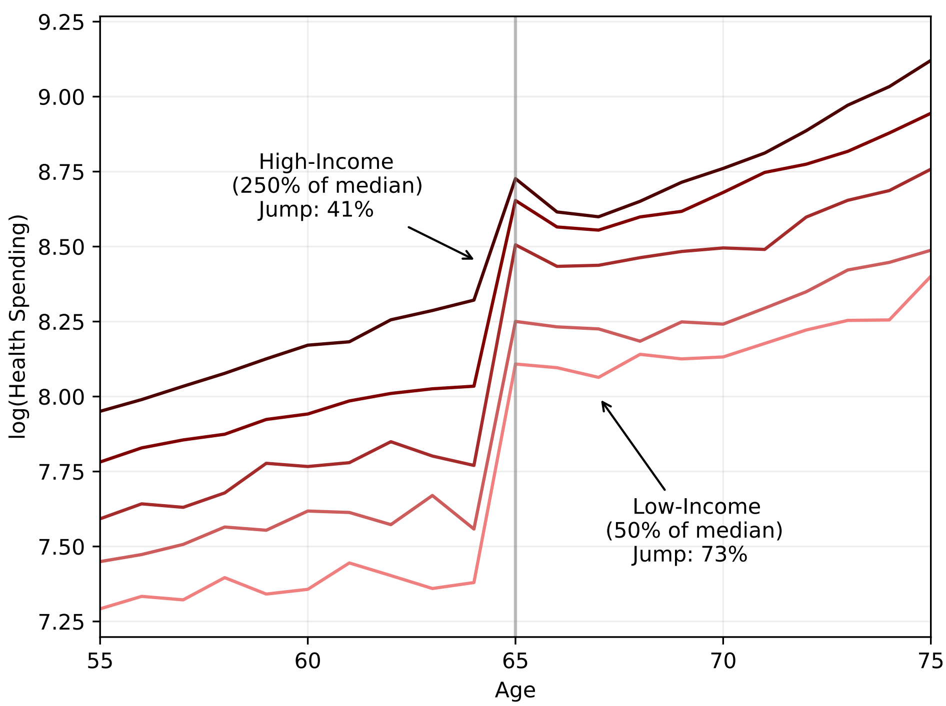


Note: This figure displays the log of average health spending. The purple line shows spending from the pre-expansion steady-state of the calibrated model while the dotted green line shows spending the from post-expansion steady-state. The jump at age 65 is estimated using a regression discontinuity design with a cutoff at age 65.

A natural first question is the extent to which expansion works to reduce delayed care. Figure 7 displays the log of average healthcare spending for individuals aged 55 to 80. The purple line displays the mean for the pre-expansion steady state while the dotted green line displays the mean for the post-expansion steady state. Expansion succeeds at reducing delayed care; the jump in spending at age 65 falls by about 12 percent (7pp) from 57 percent to 50 percent. The reduction occurs almost entirely in the three years leading up to Medicare qualification (i.e. ages 62, 63, and 64); spending at these ages is 7 percent higher after expansion. Spending by individuals younger than 62 increases by only half as much, indicating that the largest impacts of delayed care occur just around the eligibility threshold, although some shifts in timing do occur near the age 60 threshold (apparent in the figure) due to anticipation effects.

It is also clear from the figure that spending by individuals older than 65 is lower after the the expansion. Average spending by these individuals decreases by about 4 percent. Although this is smaller in relative terms than the increase for the young, it is still large in absolute terms as late-in-life expenditures are large. Of course this reduction must be weighed against the mortality channel — more older-than-65 individuals exist in the post-expansion steady-state. I defer this calculation to a later section.

Figure 8: Pre-Expansion Health Spending by Age and Income



Note: This figure displays the log of average health spending as a function of age and permanent income in the pre-expansion steady-state of the model.

The decline in delayed care, while significant, is somewhat small in magnitude — even after expansion, there remains a massive 50 percent jump in healthcare expenditure at age 65. This suggests that it is not only credit-constrained, poor individuals who are delaying care. Figure 8 demonstrates this by displaying the log of average health spending by age for individuals of various incomes in the pre-expansion steady-state. The lightest pink line displays spending for a poor individual who makes 50 percent of the median wage while the darkest red line displays the same for an individual earning 250 percent of the median (the middle three lines display 100, 150, and 200 percent respectively).

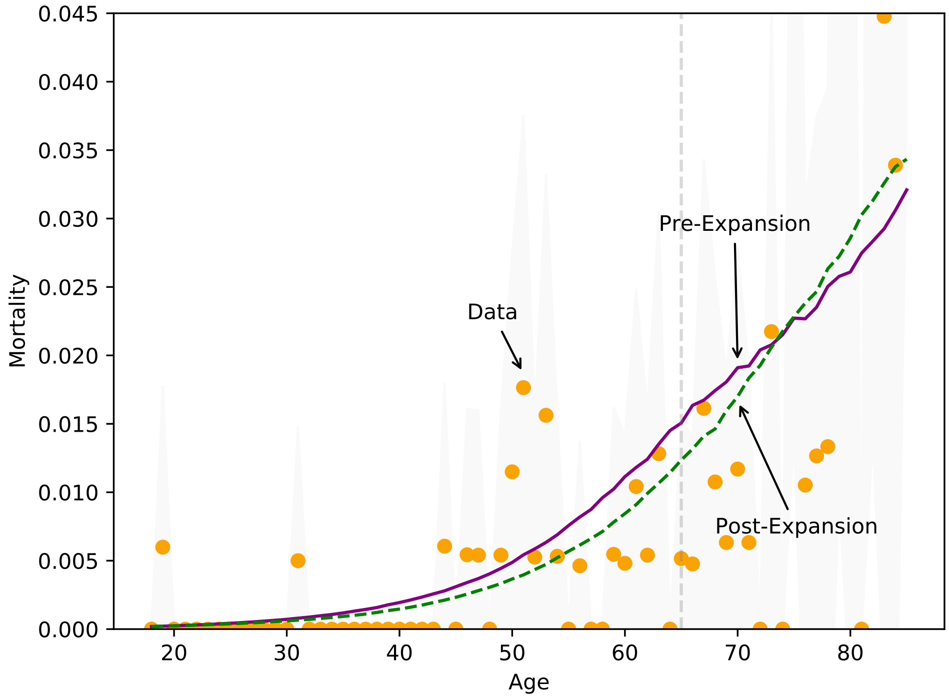
The figure confirms the fact that delayed care is a phenomenon that occurs across the income spectrum, rather than being concentrated among the poorest and most credit-constrained. Although the jump in spending at age 65 is largest for the poorest group at 73 percent, even the richest group of individuals earning multiple times the median wage exhibit a jump of 41 percent. The reason for this can be seen in Appendix Table A.3 which displays the estimated insurance plan parameters. Even for individuals with employer-provided insurance, the transition to Medicare represents a substantial drop in the out-of-pocket marginal cost of healthcare (from 28 percent to 20 percent); delaying care is the optimal response, even for the fully insured. Thus Figure 8 is a strong demonstration that it is these differences in marginal costs, rather than credit constraints, that drive delayed care. I return to this intuition in Section 6.

5.2. Reduction in Mortality

How substantial is the reduction in mortality due to less delayed care? Figure 9 displays mortality as a function of age in the pre-expansion (solid purple line) and post-expansion steady states (dotted green line) along with the data values measured in the MEPS (orange dots). The decline in mortality is substantial — recall the model is estimate to match the decline in mortality measured in Miller et al. (2021) — and is largest for individuals near the age 65 threshold, consistent with the notion that a large portion of the decrease arise from a reduction in delayed care. Looking from ages 60 to 70, average annual mortality declines by about 0.3 percentage points resulting in a 2.6 percentage point higher likelihood of survival to age 70 conditional on reaching age 60.

Around age 75, the post-expansion mortality rate is actually higher than the rate before expansion. This the result of a selection effect that occurs due to the fact that individuals who experience bad series of health shocks surviving longer. For the same reason, average health of those older than 75 counter-intuitively drops by about 0.1 (one-quarter of a deficit) after the expansion. While these changes appear malignant at first glance, they are both actually the result of improved longevity and represent increases in welfare.

Figure 9: Model Implied Mortality Pre- and Post- Expansion



Note: This figure displays average mortality as a function of age. The purple line shows mortality from the pre-expansion steady-state of the calibrated model while the dotted green line shows mortality the from post-expansion steady-state. The orange dots display mortality as measured in 2018 MEPS data with the 95 percent confidence interval shaded.

5.3. Delayed Care and the Cost of Expansion

The reduction of delayed care due to public health insurance expansion impacts the cost of expansion through both the cost-effectiveness channel (treating disease earlier is more cost effective) and the mortality channel (treating disease early saves lives and increases required Medicare outlays). Table 1 summarizes the cost of expansion in the baseline model, as well as auxiliary models aimed at isolating the impacts of these two channels.

Column 1 of Table 1 displays the increase in Medicaid coverage and spending on Medicaid and Medicare due to Medicaid expansion in the baseline model. Unsurprisingly, coverage increases by 9 percent, requiring an increase in Medicaid (public insurance for the poor) spending of about 0.3 percent of GDP. More interestingly, Medicare (public insurance for the old) expenditures decrease by 0.02 percent of GDP in the new post-expansion steady-state. As a ratio, this is implies that Medicare costs decline by around \$8 for every \$100 spent on expanding Medicaid. The fact Medicare costs decrease, rather than increase, indicates that the improvements in cost-effectiveness due

Table 1: Changes in Healthcare Spending as a Result of Expansion

Variable	(1) Baseline	(2) $\tilde{\pi}$, Fixed Prices	(3) $\tilde{\pi}$, New Prices
Medicaid Coverage (% of Pop.)	+9%	+9%	+9%
Medicaid Spending (% of GDP)	+0.32%	+0.56%	+0.56%
Medicare Spending (% of GDP)	-0.02%	-0.08%	-0.08%
Medicare Savings per $100	$8	$15	$14

Note: This table displays the increase in Medicaid coverage (in percentage of the population), increases in Medicaid and Medicare spending (in percentages of GDP), and savings on Medicare per $100 spent on Medicaid (in USD) implied by the estimated model. Column 1 displays these changes for the baseline model while Columns 2 and 3 display these changes in an auxiliary model with exogenous mortality (explained in detail in the text) where prices are fixed to their pre-expansion steady-state values ("Fixed Prices") or allowed to adjust to their post-expansion values ("New Prices").

to lower delayed care outweigh the increases in expenditure due to lower mortality (i.e. the cost-effectiveness channel is quantitatively larger than the mortality channel).

This moderate reduction in Medicare outlays, however, does not distinguish between the case where both the cost-effectiveness and mortality channels are themselves moderately sizes and the case where the two channels are quantitatively large but move in opposite directions, resulting in only a moderate net change. To examine the two channels in isolation, I introduce an auxiliary model in which the mortality function $\pi(a,h)$ is replaced by an alternative function $\tilde{\pi}(a)$ that does not depend on health. This mechanically eliminates the mortality channel as improvements in health no longer reduce mortality, leaving only the cost-effectiveness channel. However, it also eliminates the primary incentive for individuals to invest in their health; to maintain this incentive, the auxiliary model also adds a health-dependent term to individuals' utility function. I choose this term precisely so that the optimal policy function for medical spending is identical to that of the baseline model, ensuring that any differences in outcomes between the two models are due to the impact of $\tilde{\pi}$ on the distribution of households rather than change in household behavior.[18] I choose $\tilde{\pi}$ so that average mortality for each age a in the auxiliary model is identical to average (pre-expansion) mortality in the baseline model.

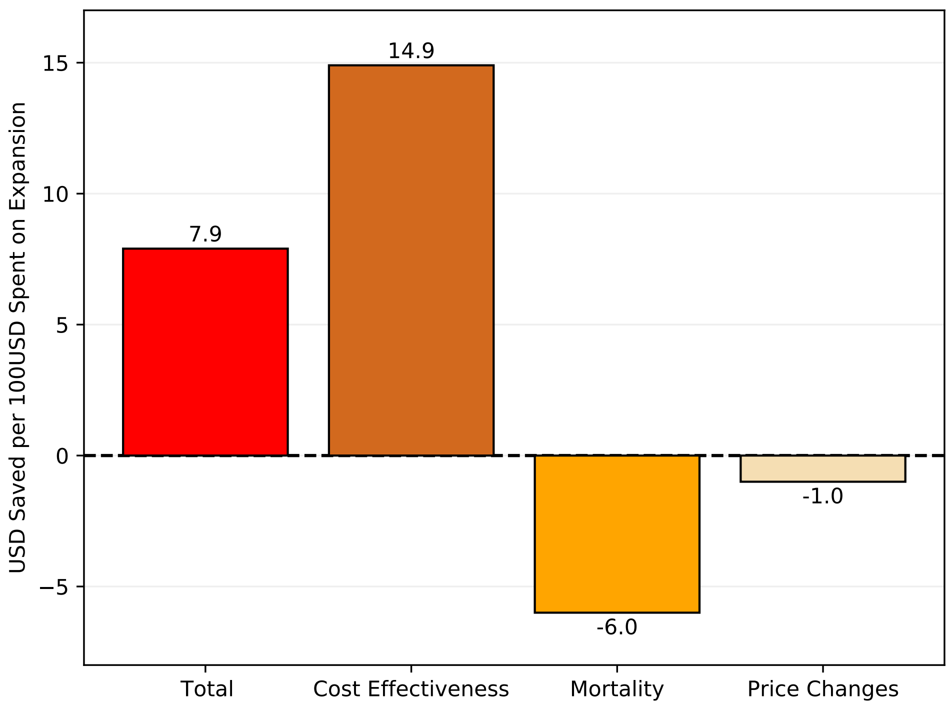
Columns 2 and 3 of Table 1 display the same results for this auxiliary model with ex-

[18] Appendix B contains a brief proof that such a function exists for any parameterization of the model.

ogenous mortality. In order to assess the impact of price and tax changes, I further separate these results into the case where these values are fixed to their pre-expansion values (Column 2) and the case where these values are allowed to adjust (Column 3). With exogenous mortality and fixed prices, Medicare costs decrease by $15 for every $100 spent on expansion. Because this model mechanically shuts off the mortality channel as well as price adjustments, I interpret this to be purely the impact of the cost-effectiveness channel.

Allowing prices to adjust (moving from Column 2 to Column 3) shrinks these savings by $1 (per$100 spent), almost entirely resulting from higher taxes which reduce incomes and very slightly increase delayed care.[19] Then the difference between Column 3 and Column 1 (whose underlying models differ only in the dependence of mortality on health) yield the impact of the mortality channel, which is a $6 increase.

Figure 10: Medicare Savings per $100 Spent on Expansion



Note: This figure displays the total savings on Medicare per $100 spent on Medicaid expansion, as well as the contributions of the cost effectiveness, mortality, and price change channels. Refer to the text for details on how this decomposition is constructed.

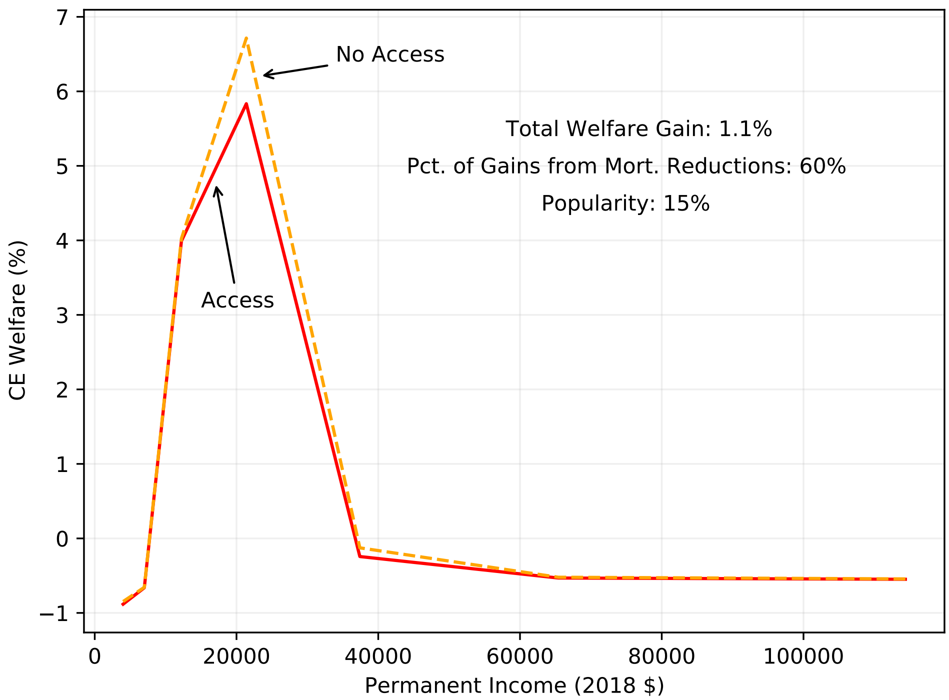
For clarity of exposition, Figure 10 displays the total decline in Medicare expenditures per $100 spent on expansion as well as the contributions of each component. In

[19] The relative price of healthcare and the price of insurance change only very slightly, stemming from the fact that spending on expansion is small relative to total healthcare spending.

total, expenditures decrease by about \$8. This is the net result of a substantial decline of \$15 due to the cost-effectiveness channel and an increase of \$6 due to the mortality channel. Overall although both channels are quantitatively important, the cost-effectiveness channel quantitatively dominates. Finally, changes in prices and taxes have little impact on delayed care.

5.4. Welfare

Figure 11: Welfare for Newborn Individuals by Permanent Income



Note: This figure displays the change in welfare, measured in consumption-equivalence, for newborn individuals (i.e. age 18) as a result of public health insurance expansion as a function of an individual's permanent income. The solid red line displays welfare for individuals born with access to employer-provided insurance while the dotted orange line displays the same for individuals born without access.

Figure 11 displays the consumption equivalent welfare change for newly born (i.e. age 18) individuals as a function of the permanent income and whether they are born with (solid red line) or without (dashed orange line) access to employer-provided insurance. This figure highlights the heavily redistributive nature of the policy. A small portion of the population, those who gain access to Medicaid, experiences very large gains on the order of 4 to 7 percent of consumption. The middle and top of the income distribution experience small losses ranging from 0 to 0.5 percent of consumption as a result of paying higher taxes and being extremely unlikely to ever receive Medicaid.

This is reflected in the popularity of the policy; only 15 percent of the population experience a welfare gain from expansion while the remaining 85 percent experience welfare losses. Still, the gains to the very poorest are sufficiently large to outweigh these disperse losses and total welfare increases by 1.1 percent.

In order to separate these welfare gains into the impact of higher consumption and the impact of longer life expectancy, I set the preference parameter governing the utility gained from living to see each period $\bar{u}$ to zero and recompute the welfare gains (without allowing reoptimization). This removes the direct utility gains from longevity and leaves behind only the gains from higher consumption.[20] The difference between this value (0.4 percent) and the total gains (1.1 percent) describes the increase in welfare that comes directly from improved longevity. Overall, these gains account for 60 percent of the total welfare gains. Unsurprisingly, there is substantial ex-post heterogeneity in who collects these benefits. Appendix Figure A.4 displays the ex-post change in welfare based on an individual's health status at age 40 (a function of their early-in-life health shocks). Individuals who end up with a health of 0.5 at age 40 experience gains as large of 15 percent of consumption while those who end up with full health experience gains of only around 2 percent.

6. A Policy to Minimize Delayed Care

Given that public insurance expansion results in only a small reduction in delayed care, it is natural to think about how expansion compares to a policy designed explicitly to minimize or eliminate the incentives to delay care. The results of Section 5.1 indicate that delayed care is mostly a response to the sudden decline in the marginal cost of healthcare that occurs at age 65 rather than a result of credit-constraints. A natural policy to consider, then, is one that slowly phases in public insurance for the elderly to avoid such a sudden jump in costs. It is worth noting policies along these lines do exist; for example, Japan's public insurance scheme featured a coinsurance rate that declined from 30 percent for the young to 10 percent for the elderly (see McGrattan et al. 2018, for a useful discussion) — although rates for the elderly have since been raised for budgetary reasons.

To accomplish this slow phase-in, I consider a policy in which the government pays for a portion of near-old individuals' pre-insurance healthcare expenditure while the remaining expenditure is covered by insurance as before. The percentage of spending covered by government varies with an individual's age and insurance status in order to smoothly transition from an individual's marginal cost of care under their private in-

[20] More precisely, the gains from higher consumption *and* higher probabilities of living to see that consumption.

surance plan to their marginal cost under Medicare. In particular, I focus on individuals ages 61 - 64 so that each year closes 20 percent of the gap between private insurance and Medicaid.

As a simple example, consider an age 61 individual with marketplace insurance. Absent any policy they must pay for 38 percent of their medical expenditure. At age 65, this will drop to 20 percent. To close 20 percent of this gap (or 40 percent for age 62, 60 percent for age 63, etc.), they should face a marginal cost of about 34 percent. This is accomplished if the government first pays for 9 percent of the individual's expenditure and leaves the rest to be split between the individual and their insurance.[21] The same calculation can be made for every age 61 to 64 and every insurance plan to ensure that the marginal cost of healthcare smoothly approaches Medicare's 20 percent at age 65 for all individuals. To prevent this additional government coverage distorting individuals' insurance choices, I also reduce the community-rated price ratio (which in practice determines the price) by the same percent. In this way, although expected insurance outlays for the example individual decline by 9 percent, the price of insurance also declines by 9 percent leaving the relative benefits (i.e. the loading factor) unchanged.[22]

6.1. Results of the Phase-In Policy

I begin in the pre-expansion steady-state of the model and introduce the phase-in policy. Figure 12 displays the log of average healthcare spending as a function of age in the model without (purple solid line) and with (dotted green line) the phase-in policy. It is clear from the figure that the phase-in policy effectively eliminates delayed care — the jump in spending at age 65 disappears as a result of both higher spending from ages 61 to 64 and lower spending immediately after qualifying for Medicare at age 65.

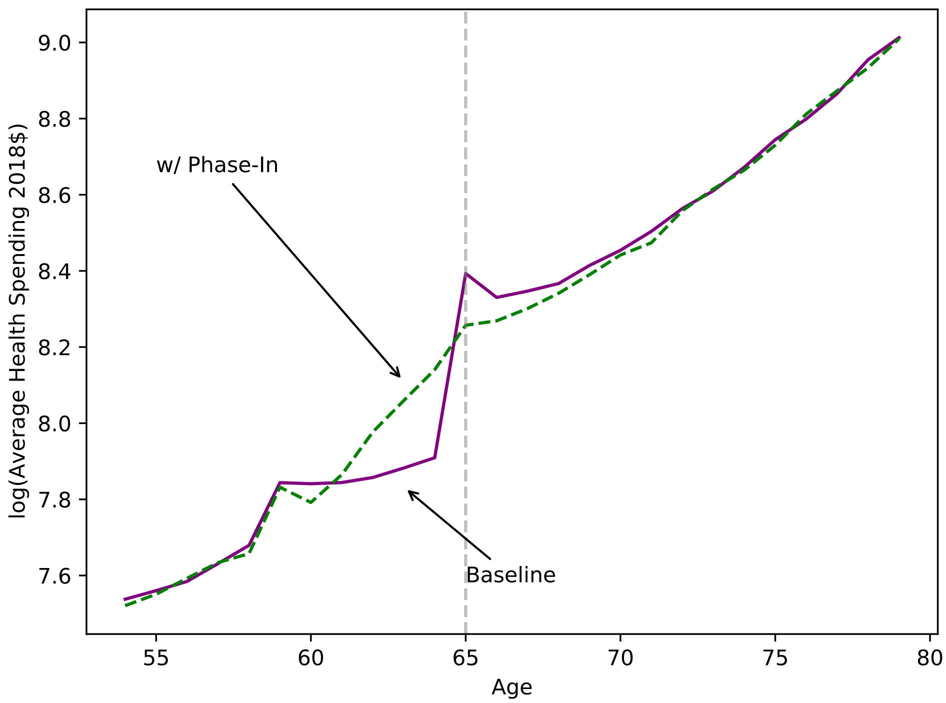
Spending for individuals between the ages of 61 and 64 increases dramatically by 16 percent on average. Although this may be unsurprising as individuals' marginal costs of healthcare are lower (relative to the baseline) during these years, it is complemented by a 6.5 percent decrease in spending between the ages of 65 and 70 — including a decline of 16 percent at age 65 specifically — which provides evidence that a large portion of this increase represents care that would have been delayed until age 65 otherwise.

As one would expect, mortality also declines substantially as a result of the intro-

[21] $(1 - 0.09) \times 0.38 = 0.34$

[22] The fact that the uninsurance "plan" has an undefined loading factor (zero divided by zero) presents one technical complication. For simplicity, I keep the price of uninsurance at zero. Although this does mean that the introduction of the smoothing policy leads to a distortion in the value of the insurance plan, this seems to have minimal quantitative impact. An alternative approach would be to increase the price of the uninsurance plan to match the loading factor of the marketplace plan. This would shrink distortions even further but leads to substantial computational complications due to the borrowing constraint.

Figure 12: Model Implied Delayed Care With and Without Phase-In



Note: This figure displays the log of average health spending. The purple line shows spending from the pre-expansion steady-state of the calibrated model while the dotted green line shows spending after the phase-in policy is implemented.

duction of the phase-in policy. Average annual mortality among those aged 60 to 70 declines by a little over 0.03 percentage points. Intuitively, the magnitude of this impact peaks right at the age 65 threshold where mortality declines by almost 0.1 percentage points. As a result of these declines, the probability of surviving to age 70 conditional on reaching 60 increases by 0.3 percentage points. This is roughly an order of magnitude less than the decline resulting from the expansion of public insurance coverage examined in Section 5. Cost-effectiveness (addressed below) aside, this highlights the difference in targeting between the two policies. Although the phase-in policy impacts the poor and uninsured, it mostly operates through small adjustments for those covered by employer-provided healthcare who make up the majority of the population (especially when weighting by healthcare spending) and are relatively unlikely to die before age 70 even at baseline. Thus even large reductions in delayed care lead to smaller impacts on mortality. In contrast, the impact of Medicaid expansion is exclusively targeted towards the poor and largest for the uninsured, both groups with substantial scope for reductions in mortality.

Implementing the phase-in policy carries an annual cost of about 0.05 percent of

GDP (10 billion USD); this is the amount that the new program spends each year. Of course, as was the case with public insurance expansion, this increase in outlays is offset to some extent by a reduction in Medicare spending. In the case of the phase-in policy, this offset is extremely large; for every $100 spent on the policy, Medicare costs decrease by $64. As a result, the net cost of the policy is only one-third (3.6 billion USD) of what would expect by looking at the outlays.

Although the magnitude of the effect is smaller, the phase-in policy is a relatively cost-effective way to reduce overall mortality among the near-old in the sense that it saves many lives per (net) dollar spent. As a point of reference, the expansion policy carries a net fiscal cost of about 55 billion USD and reduces overall mortality for individuals between the ages of 60 and 70 by 3 percentage points, resulting in a cost of about 18 billion USD per percentage point. The phase-in policy achieves a reduction of 0.3 percentage points at a net cost of 3.6 billion USD, resulting in a cost of 12 billion USD per percentage point. Thus the phase-in policy is roughly 50 percent more cost effective than the expansion policy.

These results suggest that while the public health insurance expansion does reduce the extent of delayed care, it does so fairly inefficiently. The phase-in policy achieves a more significant reduction in delayed care and, as a result, is substantially more cost-effective in reducing mortality for the near-old. Of course, the phase-in policy is fairly small in magnitude and cannot be meaningfully expanded in scope. Additionally, the expansion policy is substantially more redistributive and carries large benefits even for individuals far younger than the age 65 threshold. Thus the phase-in policy should not be viewed as a substitute for public insurance expansion and merely acts as a benchmark by which to measure its impact on delayed care.

7. Conclusion

Evidence suggests that a substantial number of U.S. individuals delay healthcare until they receive health insurance through Medicare at age 65. This paper provides a quantitative model to explain this fact and analyze the extent to which public health insurance expansion can reduce individual incentives to delay care. A key question is whether reductions in delayed care lead to cost savings, due to earlier care being more effective than late care, or lead to cost increases due to reductions in mortality increasing the population share of adults over the age of 65 who are covered by Medicare. To discipline these channels, I use quasi-experimental results from the health literature. In particular, the delayed care channel is disciplined using the jump in healthcare consumption at age 65 from the regression discontinuity of Card et al. (2008), and the decline in mortality is disciplined using the 2014 ACA Medicaid expansion diff-in-diff results from Miller et al. (2021).

My results suggest that the cost-saving delayed care channel outweighs the cost-increasing mortality channel. Expanding Medicaid increases spending by those approaching the Medicare qualification threshold (i.e. ages 62-64) by 7 percent, decreases spending by those covered by Medicare about about 4 percent, and substantially reduces mortality for the near-old. Overall, the model predicts that Medicare outlays decrease by $8 for every $100 spent on Medicaid expansion. This decline is the net result of a $15 decrease due less delayed care and a $6 increase due to higher mortality (plus a $1 increase due to price changes). Both channels are substantial and must be accounted for when evaluating the impact of insurance expansion policies.

Despite the increase in insurance coverage, there remains a substantial amount of delayed care even after insurance access is expanded. This stems from the fact that delaying care is a substantial phenomenon at all income levels, not just the poor and credit-constrained. A phase-in policy that gradually introduces insurance coverage for the elderly completely eliminates delayed care and leads to substantial reductions in Medicare costs equal to $64 per $100 spent on the policy. As a result, this policy is substantially more cost-effective than the baseline expansion policy, reducing mortality by 1 percentage point per $12 billion spent (compared to $18 billion in the baseline) although its scope is more modest.

One key limitation of this analysis is that it assumes that all medical spending is productive on the margin and, thus, that all of the observed increase in healthcare consumption at age 65 represents increased investment into individual health. If, instead, increases in consumption at age 65 are a mix of necessary, productive care and of physicians prescribing unnecessary, unproductive care to exploit Medicare reimbursement rules, my analysis will overestimate the extent to which individuals are delaying important healthcare. For example, Empirical work has argued that some prescribed treatments are overused and have very little or even negative value (Kowalski 2021, is one example for the case of mammograms). It is unclear how this would translate into my central results; such a phenomenon would mean that the model simultaneously over-predicts the cost-savings due to the reduction in delayed care but under-predicts the reduction in mortality and the resulting cost increase, pushing in opposite directions on the baseline number of $8 saved per $100 spent Further analysis that explicitly models the decision problem of physicians and leverages solid evidence on the extent of over-treatment would be valuable.

References

Aizawa, N. & Fu, C. (2020), Interaction of the labor market and the health insurance system: Employer-sponsored, individual, and public insurance, Technical report, National Bureau of Economic Research.

Benjamin, E. J., Virani, S. S., Callaway, C. W., Chamberlain, A. M., Chang, A. R., Cheng, S., Chiuve, S. E., Cushman, M., Delling, F. N., Deo, R. et al. (2018), 'Heart disease and stroke statistics - 2018 update: a report from the american heart association', *Circulation* **137**(12), e67–e492.

Card, D., Dobkin, C. & Maestas, N. (2008), 'The impact of nearly universal insurance coverage on health care utilization: Evidence from medicare', *American Economic Review* **98**(5), 2242–58.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.98.5.2242*

CDC (2009), 'Percent of u.s. adults 55 and over with chronic conditions'. [Online; accessed 17-October-2023].
URL: *"https://www.cdc.gov/nchs/data/health_policy/adult_chronic_conditions.pdf"*

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A. & Cutler, D. (2016), 'The association between income and life expectancy in the united states, 2001-2014', *Jama* **315**(16), 1750–1766.

CMS (2014), '2014 actuarial report on the financial outlook for medicaid'.

CMS (2018), '2018 actuarial report on the financial outlook for medicaid'.

Cohen, S. B. (2003), 'Design strategies and innovations in the medical expenditure panel survey', *Medical care* pp. III5–III12.

Cole, H. L., Kim, S. & Krueger, D. (2019), 'Analysing the effects of insuring health risks: On the trade-off between short-run insurance benefits versus long-run incentive costs', *The Review of Economic Studies* **86**(3), 1123–1169.

Córdoba, J. C. & Ripoll, M. (2017), 'Risk aversion and the value of life', *The Review of Economic Studies* **84**(4), 1472–1509.

De Nardi, M., French, E. & Jones, J. B. (2010), 'Why do the elderly save? the role of medical expenses', *Journal of political economy* **118**(1), 39–75.

De Nardi, M., French, E. & Jones, J. B. (2016), 'Medicaid insurance in old age', *American Economic Review* **106**(11), 3480–3520.

Fang, H. & Krueger, D. (2021), The affordable care act after a decade: Its impact on the labor market and the macro economy, Technical report, National Bureau of Economic Research.

Finkelstein, A. (2007), 'The aggregate effects of health insurance: Evidence from the introduction of medicare', *The quarterly journal of economics* **122**(1), 1–37.

Floden, M. & Lindé, J. (2001), 'Idiosyncratic risk in the united states and sweden: Is there a role for government insurance?', *Review of Economic dynamics* **4**(2), 406–437.

Fukuda, H. & Mizobe, M. (2017), 'Impact of nonadherence on complication risks and healthcare costs in patients newly-diagnosed with diabetes', *Diabetes research and clinical practice* **123**, 55–62.

Gehi, A. K., Ali, S., Na, B. & Whooley, M. A. (2007), 'Self-reported medication adherence and cardiovascular events in patients with stable coronary heart disease: the heart and soul study', *Archives of internal medicine* **167**(16), 1798–1803.

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

Heathcote, J., Storesletten, K. & Violante, G. L. (2017), 'Optimal tax progressivity: An analytical framework', *The Quarterly Journal of Economics* **132**(4), 1693–1754.

Herkert, D., Vijayakumar, P., Luo, J., Schwartz, J. I., Rabin, T. L., DeFilippo, E. & Lipska, K. J. (2019), 'Cost-related insulin underuse among patients with diabetes', *JAMA internal medicine* **179**(1), 112–114.

Hosseini, R., Kopecky, K. A. & Zhao, K. (2021), 'The evolution of health over the life cycle', *Review of Economic Dynamics* .

Jung, J. & Tran, C. (2016), 'Market inefficiency, insurance mandate and welfare: Us health care reform 2010', *Review of Economic Dynamics* **20**, 132–159.

Kaplan, G., Moll, B. & Violante, G. L. (2018), 'Monetary policy according to hank', *American Economic Review* **108**(3), 697–743.

Kaplan, G. & Violante, G. L. (2010), 'How much consumption insurance beyond self-insurance?', *American Economic Journal: Macroeconomics* **2**(4), 53–87.

Kopecky, K. A. & Koreshkova, T. (2014), 'The impact of medical and nursing home expenses on savings', *American Economic Journal: Macroeconomics* **6**(3), 29–72.

Kowalski, A. E. (2021), 'Mammograms and mortality: How has the evidence evolved?', *Journal of Economic Perspectives* **35**(2), 119–40.

Lagakos, D., Moll, B., Porzio, T., Qian, N. & Schoellman, T. (2018), 'Life cycle wage growth across countries', *Journal of Political Economy* **126**(2), 797–849.

McGrattan, E. R., Miyachi, K. & Peralta-Alva, A. (2018), *On financing retirement, health, and long-term care in Japan*, International Monetary Fund.

McKay, A. & Wieland, J. F. (2019), Lumpy durable consumption demand and the limited ammunition of monetary policy, Technical report, National Bureau of Economic Research.

McWilliams, J. M., Zaslavsky, A. M., Meara, E. & Ayanian, J. Z. (2003), 'Impact of medicare coverage on basic clinical services for previously uninsured adults', *Jama* **290**(6), 757–764.

Miller, S., Johnson, N. & Wherry, L. R. (2021), 'Medicaid and mortality: new evidence from linked survey and administrative data', *The Quarterly Journal of Economics* .

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

Ozkan, S. (2014), 'Preventive vs. curative medicine: A macroeconomic analysis of health care over the life cycle', *Unpublished. https://sites. google. com/site/serdarozkan/Ozkan_2014. pdf* .

Pashchenko, S. & Porapakkarm, P. (2013), 'Quantitative analysis of health insurance reform: Separating regulation from redistribution', *Review of Economic Dynamics* **16**(3), 383–404.

Patel, D. C., He, H., Berry, M. F., Yang, C.-F. J., Trope, W. L., Wang, Y., Lui, N. S., Liou, D. Z., Backhus, L. M. & Shrager, J. B. (2021), 'Cancer diagnoses and survival rise as 65-year-olds become medicare-eligible', *Cancer* .

Sommers, J. P. (2007), 'Methodology report #17: Additional imputations of employer information for the insurance component of the medical expenditure panel survey since 1996', https://meps.ahrq.gov/data_files/publications/mr17/mr17.shtml. Accessed: 2010-09-30.

Spitzer, S. & Weber, D. (2019), 'Reporting biases in self-assessed physical and cognitive health status of older europeans', *PLoS One* **14**(10), e0223526.

SSA (2023), 'Actuarial life table 2020 (2023 tr)'. [Online; accessed 27-September-2023]. **URL:** *https://www.ssa.gov/oact/STATS/table4c6.html*

Trabandt, M. & Uhlig, H. (2011), 'The laffer curve revisited', *Journal of Monetary Economics* **58**(4), 305–327.

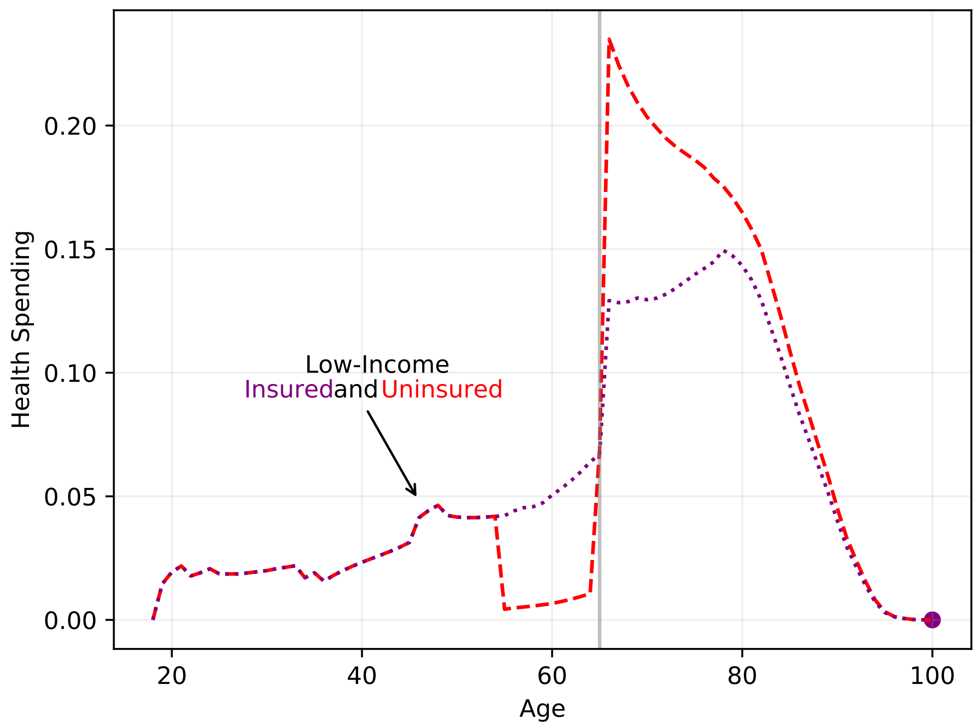
Viscusi, W. K. & Aldy, J. E. (2003), 'The value of a statistical life: a critical review of market estimates throughout the world', *Journal of risk and uncertainty* **27**(1), 5–76.

Yogo, M. (2016), 'Portfolio choice in retirement: Health risk and the demand for annuities, housing, and risky assets', *Journal of Monetary Economics* **80**, 17–34.

Zuvekas, S. H. & Olin, G. L. (2009), 'Accuracy of medicare expenditures in the medical expenditure panel survey', *INQUIRY: The Journal of Health Care Organization, Provision, and Financing* **46**(1), 92–108.

A. Additional Figures and Tables

Figure A.1: Life-cycle Health Spending for Low-Income Insured and Uninsured



Note: This figure displays life-cycle health spending for two individuals. The purple line displays the spending for an individual with low permanent income and access to employer based insurance. The red line displays spending for an individual with low permanent income and no access to employer-based insurance.

Table A.1: MEPS Medical Provider Survey Coverage

Provider	Coverage
Hospitals	100%
Office-based Physicians	100% (Medicaid + Medicare covered individuals) 75% (HMO or managed care covered individuals) 25% (remaining individuals)

Note: This table reports the percentage of medical spending covered by the Medical Provider component of the Medical Expenditure Panel Survey for 4 different categories of care. Information taken from Sommers (2007).

Table A.2: Some Summary Statistics of Investment and Emergency Spending

	All			65 or Older		
	Mean	Median	$\% > 0$	Mean	Median	$\% > 0$
Investment	\$3,847	\$924	81%	\$6,105	\$2,892	96%
Emergency	\$2501	\$0	34%	\$4041	\$0	49%

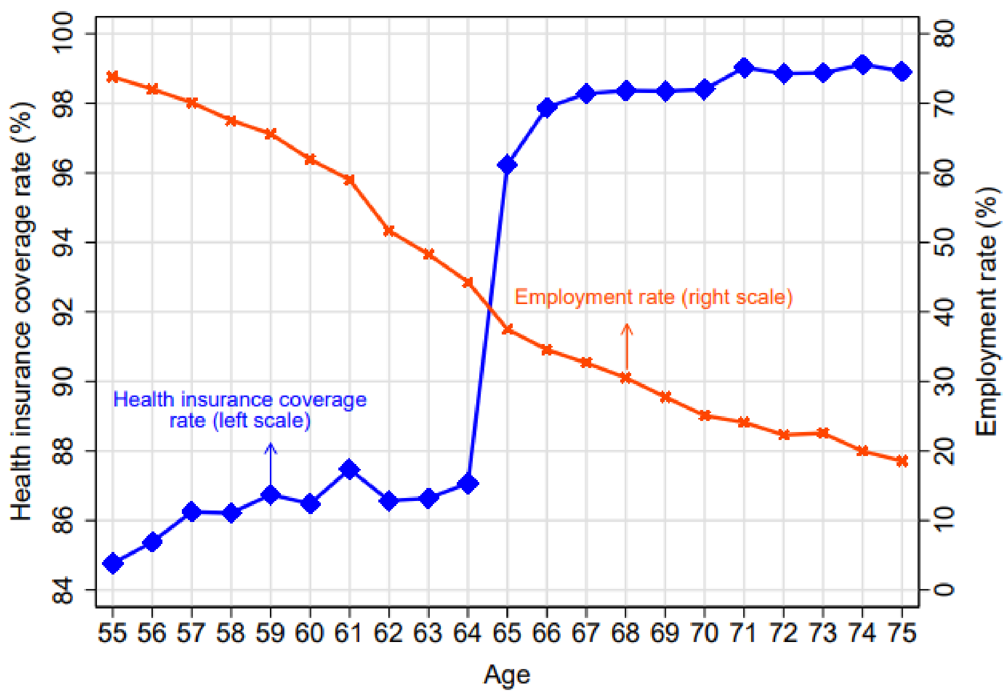
Note: This table provides some summary statistics for healthcare investment and emergency spending. Emergency spending is defined as any spending that takes place in the emergency department and investment spending is all other spending. Calculated from the Medical Expenditure Panel Survey

Table A.3: Estimated Insurance Plan Parameters

Plan	Copay Rate	Deductible	Coins. Rate	Premium
Employer	0.280	\$2,400	0.103	Deter. in Eq.
Marketplace	0.383	\$2,400	0.126	Deter. in Eq.
Medicaid	0.02	\$0	0.02	\$0
Medicare	0.20	\$1,484	0.0	\$180
Uninsured	1.0	\$0	1.0	\$0

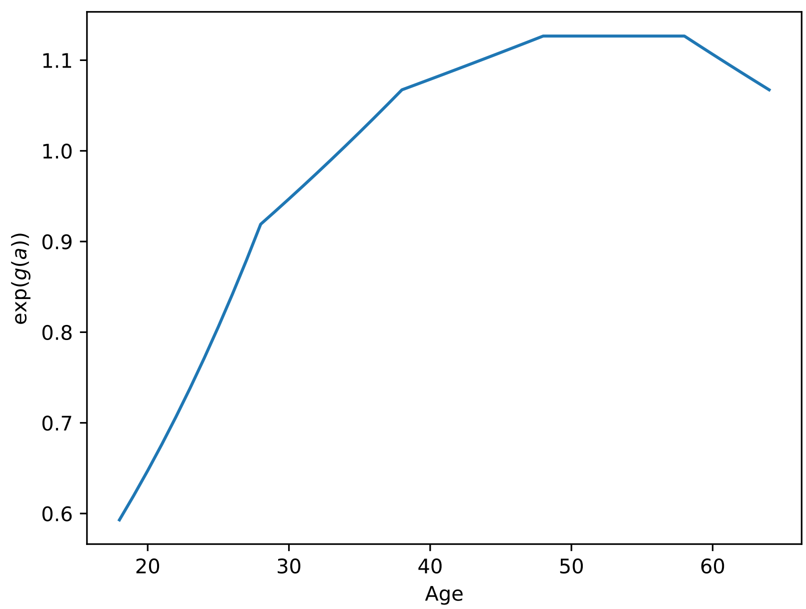
Note: This table displays the calibrated and estimated insurance plan parameter. See discussion for details

Figure A.2: Employment Rate as a Function of Age



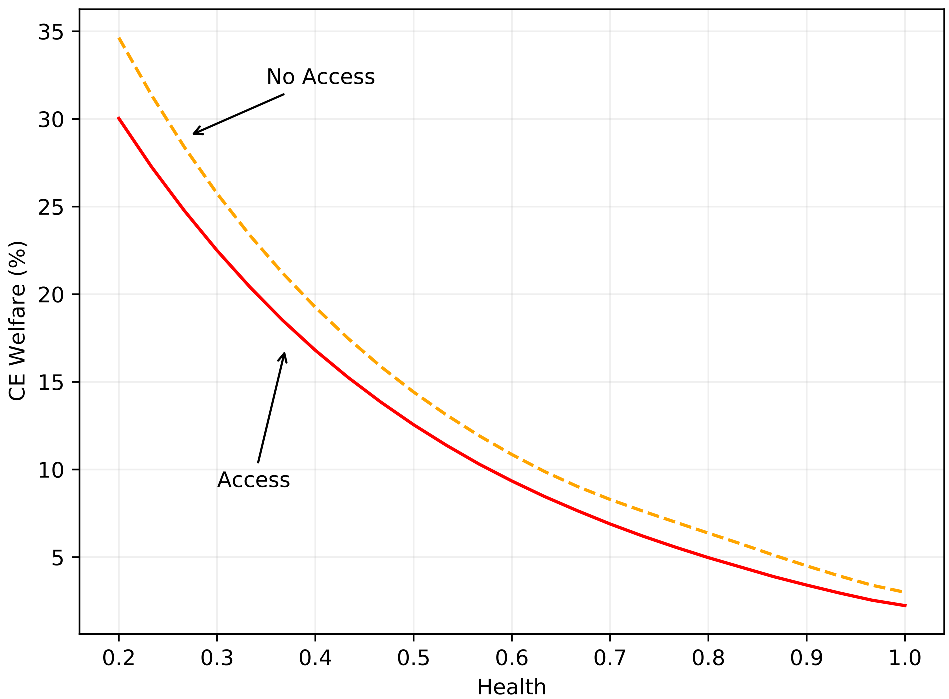
Note: This figure displays health insurance coverage rates in blue and employment rates in red as a function of age. Calculated using NHIS data from 2002 to 2012

Figure A.3: Calibrated Lifecycle Component of Income



Note: This figure displays the calibrated life-cycle component of income taken from Lagakos et al. (2018).

Figure A.4: Welfare for Newly-Covered Age 40 Individuals by Health Status



Note: This figure displays the change in welfare, measured in consumption-equivalence, for age 40 individuals as a result of public health insurance expansion as a function of individual health. The solid red line displays welfare for individuals who have access to employer-provided insurance while the dotted orange line displays the same for individuals without access.

B. Miscellaneous Model Notes

This section details various model notes and extensions.

B.1. Time Inputs into Health

A natural model extension is to allow a time input into health formation in addition to the monetary input. Exercise is the most natural interpretation of such an input, but it could also be thought of as summarizing the time that must be spent gathering information, handling administrative tasks, and traveling or waiting to receive medical care. Not only are these costs potentially important, including them is necessary to apply the model to countries where the marginal monetary cost of medical care is close-to-zero (to the consumer).

There are many possible ways to add this extension, but two natural assumptions allow the time input to be included without additional computational cost. Even stronger, these assumptions imply that the model with time inputs is isomorphic to the model without time inputs, justifying the decision to leave them out of the baseline model.

The first assumption is that the time input to health, henceforth e_t for "exercise", experiences the same decreasing returns to scale as the goods input i_t. The assumption is essential as it ensures that exercise is neither a luxury nor necessity good — without it, the "share" of exercise as an input to health would be rising or falling with income. There are many ways to ensure this, but the simplest is to add an additional term so that the health accumulation equation is now given by

$$h_{t+1} = (1 - \delta_a - \delta^x)h_t + \phi_{a,e} i_t^{\psi} + \phi_e e_t^{\psi} \tag{16}$$

where $\phi_{a,e}$ corresponds to the model parameter ϕ_a but may take a different value in the model with exercise (making this distinction facilitates discussion of the isomorphism in a few paragraphs).

The second assumption is that agents' preferences take the time input into health as a perfect substitute for the aggregate labor input — period utility takes the form $u(c_t, l_t + \alpha_e e_t)$. This assumption ensures that the marginal benefit of additional exercise, like the marginal benefit of additional medical spending, has a linear relationship with the marginal utility of consumption (through the labor-leisure condition).

Combining these two assumptions and solving for agents optimal choice of e_t yields

$$e_t^* = \left(\frac{\phi_e}{\alpha_e w_t \phi_{a,e}} \right)^{\frac{1}{1-\phi}} i_t^* \tag{17}$$

where w_t is the "wage per unit of labor aggregate" implied by the cost-minimization problem in (21). Plugging this into the health accumulation equation (16) and rearranging yields

$$h_{t+1} = (1 - \delta_a - \delta^x)h_t + \phi_{a,e}\left(1 + (\frac{1}{\alpha_e w})^{\frac{\psi}{1-\psi}}(\frac{\phi_e}{\phi_{a,e}})^{\frac{1}{1-\psi}}\right)i_t^{\psi} \tag{18}$$

The isomorphism arises from the fact that for any parameter ϕ_a from a model without a time input to health and target time-input-to-goods-input ratio γ, choosing $\phi_{a,e}$ and ϕ_e subject to

$$\phi_a = \phi_{a,e}\left(1 + (\frac{1}{\alpha_e w})^{\frac{\psi}{1-\psi}}(\frac{\phi_e}{\phi_{a,e}})^{\frac{1}{1-\psi}}\right) \tag{19}$$

$$\gamma = \left(\frac{\phi_e}{\alpha_e w_t \phi_{a,e}}\right)^{\frac{1}{1-\phi}} \tag{20}$$

generates a model that matches aggregate exercise share γ and generates identical health behaviors as the model without the time input.

B.2. Individuals' Labor Supply Decisions

Conditional on aggregate labor supply l, the labor allocation problem of the individual is

$$\begin{aligned} &\max \;\; w_m l_m + w_c l_c \\ &s.t. \;\; l = \left((1-\alpha_m)l_m^{\frac{\xi+1}{\xi}} + \alpha_m l_c^{\frac{\xi+1}{\xi}}\right)^{\frac{\xi}{\xi+1}} \end{aligned} \tag{21}$$

which yields relative labor supply to the medical section given by

$$\frac{l_m}{l_c} = \left(\frac{1-\alpha_m}{\alpha_m}\frac{w_m}{w_c}\right)^{\xi} \tag{22}$$

Thus the relative supply of labor to the medical sector $\frac{l_m}{l_c}$ exhibits a constant elasticity of ξ with respect to the relative wage $\frac{w_m}{w_c}$.

B.3. The Value Function and Recursive Competitive Equilibrium

The individual faces eight individual-level state variables. They are

1. Assets b
2. Health h
3. Age a

4. Permanent productivity z^p
5. Temporary productivity z^s
6. Insurance plan p
7. Access to employer-provided insurance e
8. Information status χ

They also face an aggregate state variable Ω describing the cross-sectional distribution of all individuals across the 8 individual-level states. The individual problem for a well-informed individual can be written recursively as in (23). $G(\Omega)$ is the perception function used by the individual to forecast the future aggregate state. The problem for a poorly-informed individual is similar but replaces the actual health-related stochastic processes $\pi(h, a)$, $\pi_x(h, a)$, $\mu(h, a)$, and $\sigma(h, a)$ with their perceived counterparts $\pi(h^*, a)$, $\pi_x(h^*, a)$, $\mu(h^*, a)$, and $\sigma(h^*, a)$for $h^* = (1 - \chi)h + \chi$.

$$
\begin{aligned}
V(b,h,a,z^p,z^s,p,e;\Omega) =& \max \quad \bar{u} + u(c,l) + \\
& + \beta\big(1-\pi(h,a)\big)\mathbb{E}[V(b',h',a+1,z^p,z^{s'},p',e';\Omega')]
\end{aligned}
$$

$$
\begin{aligned}
s.t. \;\; c + b' + p_h\chi_p(i,m) =& \big(1+r(\Omega)\big)b + T\big((w_m(\Omega)l_m + w_c(\Omega)l_c)z(z^p,z^s,a)\big) \quad \text{if } a < 65 \\
c + b' + p_h\chi_{\text{MCR}}(i,m) =& \big(1+r(\Omega)\big)b + y_{a\geq 65}(z^p) \quad \text{if } a \geq 65 \\
h' =& (1-\delta_a-\delta_x)h + \phi i^{\psi} \\
l =& \nu\big((1-\alpha_m)l_m^{\frac{\xi+1}{\xi}} + \alpha_m l_c^{\frac{\xi+1}{\xi}}\big)^{\frac{\xi}{\xi+1}} \\
p' \in& \{\text{EMP}, \text{IND}, \text{UN}, \text{MCD}\} \text{ according to eligibility} \\
\\
b' \geq& 0 \\
i \geq& 0 \\
\\
z^{s'} \sim& \rho z^s + \varepsilon, \;\; \varepsilon \sim N(0,\sigma) \\
x \sim& \begin{cases} 0 & \text{with prob. } 1-\pi_x(h,a) \\ \log N\big(\mu(h,a),\sigma(h,a)\big) & \text{with prob. } \pi_x(h,a) \end{cases} \\
e' \sim& M(e) \\
\\
\Omega' =& G(\Omega)
\end{aligned}
\tag{23}
$$

The recursive competitive equilibrium of the model consists of

a) Individual value and policy functions for both good-information and bad-information individuals given by $(V, c, i, b', l_h, l_c, p')$ and $(V_\chi, c_\chi, i_\chi, b'_\chi, l_{h,\chi}, l_{c,\chi}, p'_\chi)$

b) Firm policy functions (L_m, K_m, L_c, K_c)

c) Price functions $(r, w_m, w_c, P_{\text{EMP}}, P_{\text{IND}})$

d) Perception function G

such that

1) The value and policy functions in a) solve the individual optimization problem (23)

2) The firm policy functions solve the firm optimization problem:

$$\max \; p_m A_m L_m^{\alpha} K_m^{1-\alpha} - rK_m - w_m L_m$$
$$\max \; A_c L_c^{\alpha} K_c^{1-\alpha} - rK_c - w_c L_c$$

3) Markets clear: $\int b d\Omega = K_m + K_c$ and (6) and (7) hold

4) Perceptions are correct: $\Omega' = G(\Omega)$

B.4. Estimated Model Parameters

This subsection fills in the model estimation details missing from the discussion in 4.2. Table B.1 lists the model parameters taken from external sources or set to standard values. The persistence and variance parameters of the income process are taken from Floden & Lindé (2001). The life-cycle component of income is taken from Lagakos et al. (2018). The functional form for the tax function as well as the progressivity and level parameters are taken from Heathcote et al. (2017) who estimate this simple functional form as a continuous and smooth approximation to the US income tax schedule.

Table B.1: Externally Calibrated Parameters

Description	Parameter	Value
Discount Factor	β	0.97
Utility from (c, l)	$u(c, l)$	$\frac{1}{1-\sigma}c^{1-\sigma}\left(1-\kappa(1-\sigma)l^{1+\frac{1}{\nu}}\right)^{\sigma}$
Coefficient of Relative Risk Aversion	σ	2
Frisch Elasticity of Labor	ν	1
Disutility of Labor	κ	0.15
Income Persistence	ρ	.91
Income SD	σ	.04
Life-cycle Income	$g(a)$	See Figure A.3
Labor Share	α	0.66
Healthcare Labor Supply Elasticity	ξ	2.22
Tax Function	$T(y)$	$\lambda_\tau y^{1-\tau}$
Tax Progressivity	τ	0.181
Tax Level	λ_τ	0.90
Social Security Function	$y_{a\geq 65}(z_p)$	See discussion

Note: This table displays model parameters calibrated to standard literature values. See discussion for details on each parameter.

Table B.2 lists the parameters of the model that are estimated directly or close-to-directly from data, their values, and the data source on which they are estimated. The distribution of individual-level permanent income is chosen to be log-normal with the mean and variance parameters calibrated to match US GDP per capita and median personal income.[23] The effective loading factors $(1 - s_{\text{EMP}})\kappa_{\text{EMP}}$ and κ_{IND} are calculated as the ratio of total premiums paid over total covered costs for all individuals in the MEPS covered by employer-provided and marketplace insurance respectively. The Markov process for the availability of employer-provided insurance is chosen to match a ratio of employer-covered individuals over marketplace-covered and uninsured individuals of 3.6 as well as an annual hazard rate of losing employer-provided insurance of 7.8 percent for working-aged individuals. Both of these values are calculated directly from MEPS data. The post-expansion Medicaid productivity cutoff $\bar{z}$ is chosen so that Medicaid is offered to all individuals earning less than 138 percent of the federal poverty level for a single adult, the level prescribed by the ACA expansion of Medicaid.

The elasticity of substitution between labor supplied to the medical and consumption sectors ξ is backed-out from the quasi-experimental results of Finkelstein (2007) who uses a difference-in-difference framework exploiting pre-existing differences in elderly insurance coverage and the national implementation of Medicare to estimate the effect of Medicare on aggregate healthcare outcomes, including hospital employment and hospital payroll. She finds that Medicare increased employment by 25.6 percent and payroll by 40.1 percent in the 5-10 years following implementation (the longest-run time horizon available). The results suggest that earnings per worker increased by 11.5 percent. I treat both the increase in employment and the increase in earnings per worker as increases in the relative employment share and relative wage for healthcare yielding an elasticity of approximately 2.22 ($= \frac{.256}{.115}$).

The mortality and emergency expenditure functions are all estimated directly from individual-level data on age, health, mortality, and emergency expenditure in the MEPS. All four are estimated using OLS with individual health, age, and the square of age as regressors.[24] In the case of the variance of emergency expenditure, I construct the individual-level variance for each observation as the squared residual from the regression used to estimate mean emergency expenditure. I then regress this individual-level variance on the predictors which recovers the best linear predictor of $\mathbb{E}([Y_i - \mathbb{E}(Y_i|X_i)]^2|X_i)$ where Y_i is emergency expenditure and X_i are the predictors which is exactly the conditional variance. The procedure is very similar to performing a Breusch–Pagan test of heteroskedasticity. Columns (1), (4), (7), and (10) display the regression

[23] Using the mean and median as targets rather than a moment summarizing the tail behavior of the income distribution such as the top 1 percent income shares reflects the fact that expansion of public health insurance is a policy that mostly affects low- and middle-income individuals.

[24] Using OLS even for binary outcomes like mortality ensures that the estimated marginal benefit of health is constant (or monotonic if the square of health were included) everywhere. This is not necessarily true in logit or probit estimation, and any non-monotonicities would substantially complicate computation of the model.

Table B.2: Directly Estimated Parameters

Description	Parameter	Value	Data
Mortality Function	$\pi(h,a)$	Table B.3 Col. 1	MEPS
Emerg. Prob. Function	$\pi_x(h,a)$	Table B.3 Col. 4	
Emerg. Mean Function	$\mu(h,a)$	Table B.3 Col. 7	
Emerg. Var Function	$\sigma(h,a)$	Table B.3 Col. 10	
Percent Healthcare Labor	α_h	See discussion	ACS
Medicaid Prod. Cutoff	$\bar{z}$	0.68	Statutory
EMP availability	M	$\begin{bmatrix} .922 & .078 \\ .281 & .719 \end{bmatrix}$	MEPS
EMP Load Factor	$(1-s_{\text{EMP}})\kappa_{\text{EMP}}$	0.67	MEPS
IND Load Factor	κ_{IND}	1.30	MEPS
Insurance Plans		See Table A.3	
Perm. Income	z_p	$\log N(\mu_z, \sigma_z)$	See discussion

Note: This table displays model parameters estimates directly or close-to-directly from data as well as the data source or aggregate target. See discussion for details on each parameter.

Table B.3: Detailed Results of Mortality and Emergency Spending Regression

VARIABLES	(1) Mortality	(2) Mortality	(3) Mortality	(4) Emerg. > 0	(5) Emerg. > 0	(6) Emerg. > 0
Health	-0.0735***	-0.0675***	-0.0456***	-0.893***	-0.832***	-0.692***
	(0.0125)	(0.0123)	(0.0122)	(0.0376)	(0.0388)	(0.0427)
Age	-0.00168***	-0.00142***	-0.00176***	-0.00573***	-0.00463***	-0.00681***
	(0.000352)	(0.000336)	(0.000356)	(0.00112)	(0.00113)	(0.00112)
Age^2	1.89e-05***	1.59e-05***	2.03e-05***	4.76e-05***	3.79e-05***	6.12e-05***
	(3.93e-06)	(3.72e-06)	(3.99e-06)	(1.14e-05)	(1.16e-05)	(1.14e-05)
Observations	11,192	10,909	11,192	11,192	10,909	11,192
R-squared	0.026	0.025	0.031	0.077	0.087	0.087
Controls		YES			YES	
Self-Reported Health			YES			YES

VARIABLES	(7) log(Emerg.)	(8) log(Emerg.)	(9) log(Emerg.)	(10) Variance	(11) Variance	(12) Variance
Health	-1.019***	-1.497***	-0.845**	-0.948	-0.938	-0.0123
	(0.301)	(0.319)	(0.346)	(0.760)	(0.814)	(0.820)
Age	0.00800	-0.00401	0.00727	0.0654***	0.0712***	0.0632**
	(0.0111)	(0.0114)	(0.0111)	(0.0247)	(0.0253)	(0.0246)
Age^2	2.43e-05	0.000115	3.53e-05	-0.000610**	-0.000667***	-0.000566**
	(0.000106)	(0.000107)	(0.000106)	(0.000237)	(0.000243)	(0.000237)
Observations	2,188	2,109	2,188	2,188	2,109	2,188
R-squared	0.031	0.055	0.031	0.005	0.008	0.008
Controls		YES			YES	
Self-Reported Health			YES			YES

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: This table displays the results of regressions of various health outcomes on health, age, and controls. Columns 1 through 3 display the results of logit regression of mortality on age and health with no controls, a battery of controls, and a control for self-reported measures of health respectively. Columns 4 through 6 display the same regression with the probability of positive emergency expenditure as the outcome. Columns 7 through 9 and 10 through 12 display the results of linear regression for the outcomes of log emergency spending of the variance of log emergency spending respectively. Calculated using MEPS data from 2018.

results for the outcomes of mortality, greater than 0 emergency expenditure, mean emergency expenditure, and the variance of emergency expenditure respectively. These are the estimated functions used in the model. The remaining columns correspond to robustness checks for additional controls.

I allow the labor disutility share of healthcare labor α_h to vary as a deterministic function of permanent income in order to capture the notion that healthcare workers are disproportionately high income. To discipline this with data, I turn to the American Community Survey (ACS). I limit my sample to employed adult individuals and estimate the probability that a given individual is classified as working in the healthcare industry as a function of the log of individual income using logit regression. The predicted probabilities, denoted $P(\text{healthcare}|\log(\text{income}))$, range from about 3 percent at the bottom of the income distribution to over 15 percent at the top of the income distribution. Under the normalization that the baseline steady-state relative wage of healthcare is equal to one[25], the relative labor supply curve gives a straight-forward relationship between relative labor supply towards healthcare and α_h, allowing me to choose α_h as a function of z_p to precisely match the pattern found in the data.

Table B.4: Parameters Estimated by SMM

Moment	Model	Data	Source	Parameter
Avg. VSL	\$10 million	\$10 million	Viscusi & Aldy (2003)	$\bar{u}$
Jump in Medical Exp. at 65	See Figure 5		Card et al. (2008)	ψ
Mortality Drop from ACA	See Figure 4		Miller et al. (2021)	ϕ_a
Avg. Health Spending	\$6,220	\$6,086	MEPS	χ
Avg. Health	0.886	0.877	MEPS	δ
cov(Health, Age)	-1.11	-1.21	MEPS	ϕ_a
Emergency Health Diff.	-0.045	-0.090	MEPS	δ_e

Note: The table displays the moments targeted in the simulated method of moments estimation along with their value in both the estimated model and the data. Also displays a rough correspondence between targeted moments and model parameters. See discussion for more details.

The remaining parameters are estimated using the simulated method of moments. Table B.4 lists the targeted moments, their values in the model and data as well as the source of the data value, and their rough correspondence to model parameters. The

[25]This normalization is possible because $\frac{w_h}{w_c}$ and $\frac{p_h A_h}{A_c}$ are not separately identified in steady-state

most straightforward correspondence is that between the parameter $\bar{u}$, which governs the utility an individual received for being alive each period, and individuals' value of statistical life within the model. I use a value of \$10 million from Viscusi & Aldy (2003).

Together, the average level of healthcare spending and average level of health pin down the depreciation of health δ and the parameter describing the perceived health benefits for poorly informed individuals χ. Finally, the covariance between health and age pins down the slope parameter governing for the productivity of health spending varies with age, and the difference in average health between those who did not experience a health emergency in the last year and those who did pins down the one-time depreciation that occurs as the result of an emergency shock δ_e.