Health Insurance For Redistribution*

Angie Acquatella, Myles Wagner, and Anthony Yu[†]

November 29, 2022 JOB MARKET PAPER

Click here for most recent version

How can governments use health insurance policy and taxes to advance health equity and reduce income inequality? We derive sufficient statistics formulas for optimal health care subsidies and taxes for a social planner who cares about health inequality, in addition to income inequality. These depend on the planner's social preferences, income-specific demand elasticities of medical spending, the joint distribution of health, income, and medical spending. We revisit the RAND Health Insurance experiment to estimate the elasticities and we find that low-income individuals have less elastic demand for medical spending than high-income individuals, suggesting that the fiscal externality of public insurance depends on the socioeconomic status of individuals who receive the health care subsidy. We simulate the optimal health insurance policy and tax schedule under egalitarian and Rawlsian welfare objectives over health and income. A planner who places a high weight on the sick, relative to the healthy, chooses to provide a health insurance policy that looks like Medicaid: the optimal health care safety net eligibility threshold is 130% of the Federal Poverty Line, subsidizing 100% of medical spending for low income individuals and 70% for the rest.

^{*}This material is based upon work supported by the National Institute on Aging under Grant Number T32-AG000186, and by the National Science Foundation under Grant Number DGE 1745303.

[†]Acquatella: Postdoctoral Fellow at Boston University Questrom; email: acqua@bu.edu. Wagner: Graduate student at Harvard University Economics; email: mnwagner@g.harvard.edu. Yu: Graduate student at Harvard University Economics; email: ayu@g.harvard.edu.

1 Introduction

How can governments use taxes and transfers to advance health equity? Designing an equitable and efficient health care system has become increasingly important in light of rising health inequality: low-income individuals have poorer health, shorter lives, and have benefited less from technological progress in medicine than high-income individuals (Chetty et. al. 2016). In this paper, we consider the design of optimal taxes and transfers, focusing on the role of public health insurance as a policy tool for redistribution and under a generalized notion of equity that incorporates individual's health into their consumption utility.

Tax policy and health insurance policy are treated independently in the literature and in the policy debate. In the context of tax policy, the value of transfers across income groups is rooted in the idea that individuals with low incomes have a high marginal utility of consumption. The individual's marginal utility of consumption, however, is affected by their health, and a transfer to a sick individual would be deemed less valuable than a transfer to a healthy individual absent some social preferences. In the context of health insurance policy, the value of transfers across health states is rooted in the idea that risk averse individuals want to equate their utility across states, and transfers to sick individuals are valuable because they provide a monetary compensation for their health misfortune. Both policies entail some form of redistribution. We aim to bridge the gap between these two seemingly disjoint literatures by considering taxes and (health insurance) transfers jointly. We do so in a framework with social preferences that rationalize transfers to poorest *and* sickest individuals in society: the existence of public health insurance, recurrence of health care topics in the policy debates across the globe, and survey evidence from Stantcheva (2020) all suggest that health equity is an important component of social preferences.

When social preferences include advancing health equity, using both health insurance and taxes as a means of redistribution dominates using the income tax alone because medical spending is more informative about an individual's health than income alone. In this paper, we focus on the policy instruments most used by governments over the world; we treat insurance as a subsidy to health care consumption and derive the optimal linear tax and health care subsidy, jointly. We derive sufficient statistics formulas for the generosity of public health insurance, as well as the income eligibility threshold for the health insurance safety net (e.g. a subgroup of individuals with more generous insurance). Our formulas nest the standard optimal tax formulas, but when we incorporate the notion of health disparities, the optimal linear tax rate is higher than before. This is because there is a negative correlation between income and health state, resulting in a lower average marginal social welfare weight than in the standard model.

We then estimate the optimal policy for a planner whose welfare weights are inversely proportional to income and health, as well as a series of alternative welfare objectives. The optimal policy involves a trade-off between the fiscal externality of the policy against the welfare benefits it may provide, and the sufficient statistics approach involves calculating labor supply elasticities for the former, and structural assumptions combined with data on consumption for the latter. Our approach looks almost identical; we estimate elasticities of medical spending with respect to health insurance by socioeconomic status using data from the RAND Health Insurance experiment, which consisted of a large-scale field experiment in the 1970s that randomly assigned individuals to heterogeneously generous health insurance plans. We find that low income individuals are less responsive to health insurance incentives than high income individuals. For welfare, we make structural assumptions following the health capital literature and estimate measures of Quality of Life following Cutler et. al. (2022) and Cutler and Richardson (1999). Consistent with the health inequality literature, we find that high income individuals enjoy better self-reported health and have lower risk factors that likely lead to expensive hospitalizations (Wadhera et. al. 2020).

We combine these two empirical objects to characterize the optimal health insurance and tax policy in our sufficient statistics formulas, which are derived from the standard Saez (2001) model that is generalized in Saez and Stantcheva (2016). Our setting closely mirrors a Mirrlees (1971) framework where individuals differ in their health capital, instead of ability. We allow for the possibility that the individuals health affects their costs of labor supply, as well are their marginal utility of medical spending, and leverage the individuals optimization decisions to arrive at our optimal policy formulas. We put structure on individuals utilities by bringing in a Grossman (1972) model of health capital, where medical spending affects an individuals quality of life (QoL) and survival probability. This specification allows for two individuals with the same income and consumption bundle to have different utilities, both in the static sense and in their present discounted lifetime utility.

Health Inequality and Public Health Insurance

All else equal, a healthy individual may enjoy the opera more than an equally wealthy individual with asthma, and the individual with asthma may have less years of life to derive consumption value from their income. The so-called 'gradient of health and wealth' has been a well documented phenomenon in the literature. For instance, the individuals at the bottom of the income distribution are more likely to get hospitalized (Wadhera et al. 2020), have higher rates of smoking and obesity that result in other further complications and care expenditures (Chetty et al. 2016), and more likely to suffer from severe mental health conditions (Shields-Zeeman and Smit, 2022). In Chetty et al. (2016), the authors document striking disparities in life expectancy across the income distribution, and they find that these disparities have only gotten larger over time.

Figure 1 showcases the (static) health disparities that exist along the socioeconomic gradient. The horizontal axis contains income deciles and the vertical axis contains a QoL measure, estimated following the literature on cost-benefit analysis in health that estimates Quality Adjusted Life Years (QALYs). The dependent variable is a health metric scaled between zero and one that comes from survey data in which individuals rate their health on a scale from one to five. The self-reported

health was regressed on a large vector of health conditions, age, and demographic variables, and the predicted self-reported health values are displayed below. The disparities are evident: conditional on having below the median medical spending, disparities in QoL along the income distribution are relatively small. Among high medical spending individuals, however, the disparities in QoL are much larger: the average self-reported health of individuals in the top income decile report is 20 percent greater than the average self-reported health of individuals in the bottom decile. Beyond disparities in QoL, there are significant disparities in life expectancy across the rich and the poor: women in the top 1% live 10.1 years more than the bottom 1%, and the gap is 14.6 years for men (Chetty et. al., 2016).



Note: Medical spending classification is based on whether the individual is above or below the median medical spending within income decile. Vertical lines show 95% CIs.

Figure 1: Quality of Life Disparities by Income

These disparities in health are our primary motivation for studying optimal health insurance policy and taxes for a social objective that aims to achieve greater health equity. Figure 1 clearly depicts the sense in which medical spending in a "tag" for health. Individuals with low medical spending are relatively healthy, which implies that consumption differences from differential incomes is the primary source of inequality among these individuals. The story among individuals with high medical spending, however, is very different. Consider giving an individual from the bottom income decile a cash transfer large enough to put her in the top income decile. The expected QoL of this individual would still lie below that of an equally wealthy individual in good health. Subsidizing the medical spending of a low income individual would be better targeted than the cash transfer alone.

In other words, there *is* a redistributive role for health insurance (when equity concerns include health and income), relative to the income tax alone. When the government is free to implement a

nonlinear income tax schedule and has redistributive preferences over income alone, it can achieve second-best redistribution using only the income tax schedule, and optimal insurance is determined solely by the tradeoff between risk protection and moral hazard. This is a special case of the well-known seminal result due to Atkinson and Stiglitz (1976) on indirect taxation. However, when the government has redistributive preferences over individual characteristics other than income (e.g. health), the solution to the social planner's problem involves using both insurance and taxes as instruments for redistribution.

We also consider the role of constraints on the government's ability to set policy. The linear case we consider here attempts to capture the specific policy features that are characteristic of health insurance policy in the US and other developed nations, such as France, Italy, Germany, Australia, and Japan.¹ Mainly, we focus on health insurance policies that subsidize a share of an individuals medical bill, and on subsidies that vary only along a binary income dimension: high or low income. The policy instruments with regards to health insurance are three: the subsidy for the rich, the subsidy for the poor, and the income threshold below which the planner designates an individual to be low income. While such a policy may seem overly restrictive or simple, it is surprisingly common across the world.

Table 1 below shows the different types of health insurance policies for different countries. Health insurance subsidies also may along other dimensions, such as the health state of an individual, or age group, but income is seems to be the most common.

Related Literature

A vast literature on optimal taxation has studied taxes and transfers under redistributive concerns on income, and a separate literature on health inequality has established the fact that large health disparities exist (and prevail) along the socioeconomic gradient. We view our contribution as bridging the gap between these two literatures by generalizing the formulas from optimal tax theory to incorporate the notion of health equity. Concerns for health equity can be directly nested in the generalized marginal social welfare weights approach proposed by Saez and Stantcheva (2016). We build upon the framework in Saez and Stantcheva (2016), Saez (2001).

Our paper is also related to recent empirical work examining the value of the insurance provided through programs such as Medicaid (Finkelstein, Hendren and Luttmer, 2019) and the ACA exchanges (Finkelstein, Hendren and Shepard, 2019). These papers estimate the average value of health insurance policy to program beneficiaries, in the form of willingness to pay for insurance,

¹Our objectives with the theory are not to characterize the second-best optimal non-linear health insurance contract (as in Blomqvist 1997), nor is it designed to characterize the optimal joint non-linear tax and health insurance schedules. To characterize such schedules would involve a more theoretical approach that distances us from our empirical setting, and thus we leave for future work. We characterize rather simple linear subsidies and taxes that resemble those used in the real world.

	Health Insurance Safety Nets				
Country	For Low Income Individuals	For Sicker Individuals	For Children, Elderly, or Mothers		
Australia	\checkmark				
Canada	\checkmark				
Denmark		Т	С		
England [†]	\checkmark	D, CC			
France	\checkmark				
Germany,			С		
Israel [†]	\checkmark		С		
Italy	\checkmark	D & CC			
Japan	\checkmark	CC			
Netherlands		D & CC	С		
New Zealand	\checkmark				
Norway			С		
Singapore	\checkmark				
Sweden			C & E		
Switzerland			C & M		
Taiwan	\checkmark				
Unites States	\checkmark	D & CC	C, M, & E		
D=disabled, CC=chronic condition, T=terminally ill, C=children, E=elderly, M=mothers.					
[†] : England and Israel have national public insurance for all, though some services require					
a copay or coinsurance. These countries further subsidize care for low income individuals.					
Source: The Commonwealth Fund.					

Table 1: Public Health Insurance Across Countries

and conclude that the costs of these programs exceed the benefits. But willingness to pay measures are a function of the budget constraint of individuals, and may lead to a very different valuation of public insurance program, relative to a Grossman-type measure of health capital. In this paper, we propose a different, yet complementary approach to these papers by leveraging the correlational structure of health capital and socioeconomic status for welfare analysis.

The paper proceeds as follows. In section 2, we present model of redistributive health insurance based on common health insurance policies used around the globe. We derive sufficient statistics formulas for optimal policy and highlight the three components needed in order to calibrate the policy: elasticities, joint distribution of income, spending, and health, and the redistributive preferences of the planner. In section 3, we estimate the demand elasticities of medical spending with respect to the out-of-pocket price, recovering heterogeneous behavioral responses to health insurance subsidies by income using the RAND experimental data. We also describe the empirical joint distribution of income, medical spending, and health, using data from the Medical Expenditure Panel Survey. In section 4, we simulate the optimal policy for a planner whose welfare weights are inversely proportional to health under our main specification. We also simulate the optimal policy under alternative social welfare weight specifications, including Rawlsian and inversely proportional to income weights. In section 5, we conclude.

2 A Model for Redistributive Health Insurance

We consider a Mirrlees framework in which individuals differ in their underlying level of health, instead of ability, and jointly characterize the optimal health insurance policy and tax schedule for a social planner with a general redistributive objective. Our primary objective with the theory is to derive sufficient statistics formulas that can help inform the design public health policy, guide the direction of policy reforms, and shed light on the redistributive preferences that lie underneath public health insurance programs. The key difference between our model and standard optimal tax models is that we incorporate the notion of health capital into individual utilities. Following the Grossman model of health capital, health affects individual utility in three ways: consumption, quality of life (QoL), and life expectancy (LE).

To build some intuition around how health affects utilities (and welfare), consider a healthy versus a sick individual. The sick individual may incur higher medical expenses than the healthy individual; holding all else equal, the sick individual will have a lower level of disposable income remaining for consumption. Now suppose that health insurance could cover medical expenses so as to equalize the level of consumption across the two individuals. If these two individuals had identical consumption bundles, consisting of only opera plays and food, the sick individual may not enjoy the opera play as much as the healthy individual just by virtue of being sick. The key distinction between health and preferences is that, in the counterfactual scenario of perfect health, the sick individual would enjoy the opera just as much as the healthy individual. The differences in utility from going to the opera thus cannot be ascribed to differences preferences. Rather, health captures the state-dependent nature of utility. The concept of Quality of life (QoL) attempts to capture these differences in utility by scaling down the utility level of the sick individual.

Now further suppose there existed a medical technology (e.g. an inhaler that alleviates the symptoms of asthma), covered by insurance, so that both individuals have the same consumption and quality of life. The sick individual may live a shorter life than the healthy individual, which means that their net present value of life-time utility will be lower than that of the healthy individual. This is entirely because the healthy individual will have more periods to derive positive utility. Therefore, even if health insurance and medical technology could equalize the flow utility of individuals every period, life expectancy (LE) will capture differences in aggregate life-time utility by telling us how long a sick individual might expect to live, relative to a healthy individual.

In our model, individuals choose consumption, medical spending, and their labor supply to maximize utility, which is affected by their health capital. Consumption and medical spending will therefore encode information about an individual's underlying health condition, which is private information and unknown to the social planner. The income tax and medical spending subsidy leverage the encoded information about the unobserved type in these (optimal) choices.² To the

²This implies that individuals accurately trade-off quality of life or life expectancy gains when choosing their spending, which may not fully capture the behavioral aspect of healthcare decisions. Nonetheless, our approach serves

extent that health insurance can affect quality of life or life expectancy across the socioeconomic gradient, the role of health insurance as a policy instrument for equity becomes apparent. In our model, however, individual optimization "envelopes away" this channel.³ Nonetheless, the redistributive role of insurance leverages the relationship between health, medical spending, and income.

Model

Consider a social planner with a redistributive objective that provides public health insurance for its citizens, financed through taxes. Following Saez and Stantcheva (2016), we consider the case of a planner with generalized social marginal welfare weights, which represent the value that society puts on giving an individual an additional dollar of consumption. Formally, suppose the planner cares about each individual *i* according to a generalized marginal social weight g_i , and sets policy to maximize the weighted average of individual utilities. The planner can affect the utility of individuals, U_i through public insurance, which partially finances medical expenditures for individuals in different income groups. The planner faces a budget constraint in which aggregate subsidies to medical expenditures must equal aggregate tax revenues.

Denote a particular policy by P, where P consists of an income tax schedule and a health insurance schedule. Suppose the government assigns welfare weights to individuals based on observable characteristics of individuals valued by society, such as health state and income. Suppose that the welfare weights are defined over characteristics of individuals i at the optimal policy, P. That is, while the characteristics of individuals may be endogenous to the policy, (e.g. health state may be a function of health insurance), the social welfare weights are fixed, and defined at a particular vector of individual characteristics that correspond to the equilibrium values *at the* optimal policy, P.

The government sets policy by maximizing an aggregate objective of individual's money-metric utility, weighted by a particular set of social welfare weights. Denoting an individual by i, their health by H_i , and their income by z_i , the government chooses a policy P optimally to maximize the following welfare objective:

$$W(\tilde{P}|P) = \int_{i} g_i(H_i, z_i|P) U_i(\tilde{P}) di, \qquad (1)$$

where \tilde{P} is an arbitrary policy, $g_i(H_i, z_i|P)$ the generalized marginal social welfare weight assigned to individual *i* at the optimal policy *P*, and $U_i(\tilde{P})$ the money-metric utility of individual *i* at any arbitrary policy \tilde{P} .

as a starting point for optimal insurance design, and more closely follows the standards in the tax literature. We leave behavioral considerations open for future work.

³One could incorporate the behavioral aspects of medical decision making by following the approach in Cutler et. al. (2022), which empirically estimates the value of healthcare while allowing for the possibility that medical spending decisions may not reflect optimal choices.

Shape of the policy: Suppose that (for reasons outside of the model) the government can only choose policies of the form: $P_t = (\tau, R, \hat{z}, s_L, s_H)$, where τ denotes a linear payroll tax, R denotes the demogrant, \hat{z} denotes an income-based eligibility threshold below which individuals receive the "safety net" health care subsidy, s_L denotes the health care subsidy for low income individuals (with incomes below \hat{z} , and s_H denotes the health care subsidy for the remaining citizens.

$$T(z_i) = \underbrace{\tau z_i}_{\text{linear}} + \underbrace{R}_{\text{lump-sum}} \qquad s(m_i, z_i) = \begin{cases} \underbrace{(1 - s_L)}_{\text{coinsurance}} m_i &, z_i \leq \underbrace{\hat{z}}_{\text{Medicaid}} \\ for \text{ the poor} & \text{eligibility} \\ \underbrace{(1 - s_H)}_{\text{coinsurance}} m_i &, \text{ otherwise}, \end{cases}$$

where we assume that the health insurance subsidy is greater for the poor, $s_L \ge s_H$, and that $\tau, s_H, s_L \in [0, 1]$. The optimal policy thus consists of a quintuple $(\tau, R, s_L, s_H, \hat{z})$ that maximize the social objective stated in (1), subject to the budget constraint that total tax revenues must be greater than or equal to total health care subsidies.

We consider policies of this form because they correspond closely to our empirical setting of US health insurance, and to other countries as shown in Table 1. The question of why the US Congress or other governments across the globe chose this particular structure for the policy for health insurance is beyond the scope of the model, but we could allude political economy reasons or administration costs that make such a policy appealing. More generous health insurance subsidies for the poor, for instance, has been a major subject of political campaigns since the beginnings of Medicare in the 1970's. Health insurance billing and reimbursement is notoriously costly to administer (Dunn et al., 2021), and perhaps adding layers of complexity (e.g. insurance that varies with income more granularly) would be undesirably costly. On the tax side, a flat tax with a lump sum transfer has the advantage of greatly simplifying the optimization problem, and it could be considered weakly optimal in the sense that it can approximate the optimal non-linear tax schedule (Mankiw et al. 2009; Diamond 1998).

We take as given that the government finances health insurance through a linear payroll tax, as Medicare does. Our model accounts for the fact that taxes directly affect incentives for labor supply. However, in the interest of tractability, we abstract away from the consideration that generous health insurance subsidies also may affect labor supply. An empirical literature has investigated the question on whether generous health insurance subsidies significantly reduce incentives to work, and, while evidence is mixed, the broad conclusion is that labor supply responses to health insurance are small and negative (Kaestner et al. 2017, Kucko et al. 2017, Buchmueller et al. 2019, Dague et al. 2017, Garthwaite et al. 2014). Further, we abstract away from the consideration that taxes directly affect medical spending decisions; this is because we consider local perturbations to the tax schedule which result in marginal effects on income that do not have first order effects on

medical spending.

Individuals and Information: Consider a population of individuals indexed by i, where population size is normalized to one. Individuals in society differ in how healthy or sick they are, and their health state affects utility in three ways: sicker individuals may incur larger medical expenses, they may derive lower marginal utility from consumption (Finklestein et al., 2013), or they may life shorter lives. The health state of an individual i at time t is fully captured by their type, θ_{it} . The individual's type represents his health state or wellbeing, and it fully determines how much an individual benefits from medical spending.

Individuals know their type θ_{it} and make decisions about medical spending, consumption, and labor supply, based on their type. An individual's type is private information and not directly observed by the government, but the government can observe an individual's level of income, their medical expenditures, and some health-metric characteristics (e.g. chronic conditions, disability status, kidney failure diagnosis) that are correlated to their type θ_{it} .

Utility and Health Capital: Suppose individual i of age a_t at time t derives utility from consumption c_{it} and incurs disutility from earning income $v_i(z_{it}) \ge 0$. The flow utility of the individual is scaled by their health state, $H_{it} \in (0, 1]$, where zero corresponds to death and one corresponds to perfect health. Suppose that the health state H_{it} depends on the individual's medical spending in that period m_{it} , their type θ_{it} , and their age a_t . Let the value function of individual i at time t be the sum of their flow utility at time t plus their expected continuation value,

$$V_{i,t} = u(c_{it} - v_i(z_{it}))H_{it}(m_{it}, a_{it}, \theta_{it}) + EV_{i,t+1}(m_{it})$$

where $u(\cdot)$ is increasing, concave, and common function to all individuals. Suppose that the continuation value is their net present value of future life-time utility, which depends on their future health state, their future flow utility, their discount rate, r, and the survival probabilities $S(t + k|\cdot)$ across all future periods t + k for $k \ge 1$.

Time independence: Assume there are no savings. Individuals choose labor supply and medical spending every period to maximize their utility, and consume the entirety of their income (net of medical spending) each period. Assume that the expected survival probability at time t for any arbitrary future period depends on *only on current* medical spending m_{it} and age. Assume that individuals expect to remain in the same state as in the current period until the end of their life, so that $H_{i,t+k} = H_{it}$ for all $k \ge 1$.

$$EV_{i,t+1} = \sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^{t+k} u(c_{i,t+k} - v_i(z_{i,t+k}))S(t+k|a_{i,t+k}, m_{it})H_{it}(m_{it}, a_{it}, \theta_{it})$$

Distribution of income, medical spending, and consumption: Suppose that the distribution of θ_i is such that income z_i is distributed according to $F_z(z_i)$ over the support $z_i \in [0, Z]$, and that

medical spending m_i is correlated to z_i , so that there is a joint density $F_{zm}(m_i|z_i)$ over the support $m_i \in \mathcal{M}$. Given an income z_i and medical spending m_i , and the policy P = (T, s), individual *i* faces the budget constraint,

$$c_i + (1 - s(z_i))m_i \ge (1 - \tau)z_i + R.$$

Since the marginal utility of consumption is positive, the individual's budget constraint always holds with equality, and individuals choose (m_i, z_i) according to $-(1 - s(z_i))u_c + u_m = 0$ and $(1 - \tau) + u_z = 0$, respectively.

Aggregate budget constraint: Any optimal policy P must satisfy the resource constraint in order for the policy to be feasible. Suppose the government only has tax revenues to finance health insurance subsidies (s_L, s_H) . Given our particular policy form, the aggregate budget constraint at an arbitrary policy \tilde{P} is given by,

$$B(\tilde{P}) = \underbrace{\int_{0}^{\hat{z}} \int_{m_i \in \mathcal{M}|z_i} \tilde{s_L}m_i \, dF_{zm}(m_i|z_i) dF_z(z_i) + \int_{\hat{z}}^{Z} \int_{m_i \in \mathcal{M}|z_i} \tilde{s_H}m_i \, dF_{zm}(m_i|z_i) dF_z(z_i) - \tilde{R} + \underbrace{\left(\tilde{\tau} \int_{0}^{Z} z_i dF_z(z_i)\right)}_{\text{aggregate tax revenue}} \operatorname{aggregate medical expenditures}$$

As usual, the planner trades off the marginal welfare gains of providing insurance and transfers against the fiscal externality of such policies. At the optimal policy, the marginal welfare gains are exactly equal to the fiscal costs, which means that there are no (local) Pareto improving reforms at the optimal policy. To provide some intuition around the trade-off, consider the following thought experiment. Consider a policy reform that lowers the Medicaid health insurance subsidy, s_L , while holding all other policy instruments fixed. In particular, consider a reform that lowers the Medicaid subsidy from 100% to 75%, so that Medicaid individuals now have to pay 25% of their healthcare costs. On the welfare side, making insurance less generous presumably would decrease the utility of the Medicaid beneficiaries, but it would not directly affect the utility of the non-Medicaid beneficiaries. On the financing side, reducing the Medicaid health insurance subsidy relaxes the government's budget constraint. Reducing expenditures allows the government to redistribute excess tax revenues through the lump sum rebate R. If the original policy was optimal, this reform *cannot be welfare enhancing*. That is, the weighted aggregate utility loss for the Medicaid beneficiaries must be greater than the aggregate budget savings generated by such a policy reform.

If we denote the current policy by $P = (s_L, s_H, \hat{z}, \tau, R)$ and the reformed policy by $P' = (s'_L, s_H, \hat{z}, \tau, R)$, the aggregate welfare loss of such a policy reform is given by $W(P') - W(P) = \int_i g_i \cdot (U_i(P') - U_i(P)) di$. In our thought experiment, the utility of non-Medicaid individuals remains the same under such a policy reform, which implies that the change in welfare for these individuals is zero. The welfare effect can be decomposed into the sum of Medicaid and non-

Medicaid individuals.

$$W(P') - W(P) = \int_0^{\hat{z}} \int_{\mathcal{M}_z} g_i(U_i(s'_L|\tau, R) - U_i(s_L|\tau, R)) dF_{mz}(m_i|z_i) dF_z(z_i)$$
(Medicaid)
+
$$\underbrace{\int_{\hat{z}}^Z U_i(s_H|\tau, R) - U_i(s_H|\tau, R) dF_z(z_i)}_{=0}$$
(non-Medicaid)

If the utility function is concave in the policy instrument, then we can put bounds on the utility loss from this policy reform. The basic intuition is best conveyed graphically. Figure 2 plots the utility function of individual *i* against the health insurance subsidy *s*. The utility loss from reducing the subsidy from $s_L = 1$ to $s'_L = .75$ is the vertical distance between utility at the previous subsidy versus the reformed subsidy, ΔU_i . We can put an upper bound on the magnitude of the utility loss by taking a linear approximation of the utility function *at the reformed policy* s'_L . Concave utility implies that the linear approximation evaluated at the less generous subsidy will over-shoot the utility difference. Conversely, we could put a *lower* bound on the magnitude of the utility loss by taking a linear approximation of the utility function at the *initial policy* s_L .



Figure 2: Bounds on Utility Effects from Arbitrary Policy Reform

If the original policy was optimal, then the incremental revenue gains (budget effect) from the policy reform has to be smaller than the utility loss. The linear approximation gives us a *lower* bound on the utility loss. Therefore, we say that if the savings from lowering the Medicaid subsidy from 100% to 75% are equal to the lower bound of the aggregate utility loss, the planner weakly prefers the original policy P over the reformed policy \tilde{P} . Denote the pairwise elasticity between

 $s_L = 1$ and $s_L = .75$ by $\eta_m(s_L, \Delta)$. The optimal healthcare subsidy equates these two effects.

$$\underbrace{\int_{0}^{\hat{z}} \int_{\mathcal{M}} g_i \frac{dU_i}{ds_L} \Big|_{s_L=1} dF_{z,m}(m_i|z_i) dF_z(z_i)}_{\text{lower bound on welfare loss}} = \underbrace{(\eta_m(s_L, \Delta | \mathcal{Z}') + 1) \int_{0}^{\hat{z}} \int_{\mathcal{M}} m_i dF_{z,m}(m_i|z_i) dF_z(z_i)}_{\text{savings generated by the policy reform}}$$

The welfare bounds approach allow us to consider non-local deviations and establish conditions for "weak optimality" (in the sense that the planner could either be indifferent or strictly prefer the original policy under such conditions). While, in practice, we only consider local deviations to characterize the optimal healthcare subsidies and tax rate, we need to make non-local comparisons to establish the optimal income threshold under which individuals are eligible for Medicaid.

The Optimal Healthcare Subsidy

Following the standard approach in the optimal taxation literature, we characterize the optimal policy by looking at local deviations around the policy and characterizing the point around which marginal welfare gains or losses are zero. Suppose the planner sets policy to maximize aggregate utility at a fixed point in time, taking into account only the current distribution of $V_{i,t}$ at time t. The planner aggregates money-metric utilities, U_i , where $U_i = V_i/u'(c)$ in order to re-scale utils into dollars. The optimal policy will maximize the objective in equation (1) subject to $B(\tilde{P}) \leq 0$.

To begin, let's first consider the optimal generosity of health insurance for each of the income groups, holding the income threshold fixed. Let the elasticity of medical spending for the subset of individuals with incomes $z_i \in \mathcal{Z}'$ be $\eta(s|\mathcal{Z}')$, where the elasticity is taken with respect to the coinsurance rate (1 - s). Standard optimal tax derivations using the individuals envelope theorems for the choice m_i yield:

$$\frac{dSWF}{d(1-s')}\Big|_{z_i\in\mathcal{Z}'} = -\int\limits_{z_i\in\mathcal{Z}'} g_i H_i(m_i, a_i, \theta_i) \cdot m_i \, di + \lambda \left(\frac{s'}{1-s'}\eta(s'|\mathcal{Z}') + 1\right)$$

where λ is the multiplier on the government budget constraint. At the optimal subsidy, $\frac{dSWF}{d(1-s')} = 0$, leading the the following proposition.

Proposition 1 The optimal healthcare subsidy for individuals in the income subgroup $z_i \in \mathbb{Z}'$ is given by:

$$s' = \frac{\bar{g}_m(\mathcal{Z}') - 1}{\bar{g}_m(\mathcal{Z}') - 1 + \eta(s'|\mathcal{Z}')} \quad \text{with} \quad \bar{g}_m(\mathcal{Z}') = \frac{\int_{\mathcal{Z}'} g_i \cdot H_i \cdot m_i \, di}{\lambda \int_{\mathcal{Z}'} m_i \, di} \quad \text{and} \quad \eta(s'|\mathcal{Z}') = \frac{s \cdot \frac{d \int_{\mathcal{Z}'} m_i di}{ds}}{\int_{\mathcal{Z}'} m_i di} > 0$$

$$\tag{2}$$

The optimal linear income tax can be derived symmetrically. Let the elasticity of aggregate labor income be ξ_z . The optimal payroll tax is given by:

$$\tau = \frac{1 - \bar{g}_z}{1 - \bar{g}_z + \xi_z} \quad \text{with} \quad \bar{g}_z = \frac{\int_i g_i \cdot H_i \cdot z_i \, di}{\lambda \int_i z_i \, di} \quad \text{and} \quad \xi_z = \frac{(1 - \tau)}{\int_i z_i di} \frac{d \int_i z_i di}{d(1 - \tau)} > 0 \quad (3)$$

Now we turn to the optimal income threshold. Given a fixed income threshold, the optimal healthcare subsidies are given by equation (2), but the optimal subsidies differ for different income thresholds. The optimal healthcare subsidies (and tax) determine aggregate welfare at either side of the threshold, but there is no reason to believe that the two healthcare subsidies will be locally proximate to each other. Let's suppose that \hat{z} is optimal and that the optimal health insurance subsidy for low income individuals, s_L , is more generous than the optimal health insurance subsidy for high income individuals, s_H (at this \hat{z}). Since this income threshold is optimal, the welfare gains from moving individuals into the low income group with more generous insurance (or conversely, moving them out of the generous insurance program) should not exceed the incremental costs from such a policy reform. At most, the welfare gains of such a policy reform would be equal to the upper bound of the linear approximation. Since \hat{z} was optimal, it must be the case that the incremental costs of raising \hat{z} are at least as large as the upper bound of the welfare gains (if not larger).

Proposition 2 The optimal "low-income eligibility" threshold \hat{z} is characterized by by:

$$\bar{g}_m(\hat{z}) = (\eta_m(s_H, s_L - s_H | \hat{z}) + 1) \qquad \text{with} \qquad \bar{g}_m(\hat{z}) = \frac{\int_{\mathcal{M}} g_i H_i \cdot m_i \, dF_{m,z}(m_i | \hat{z})}{\lambda \int_{\mathcal{M}} m_i dF_{z,m}(m_i | \hat{z})},$$

and where $\eta_m(s_H, s_L - s_H | \hat{z})$ denotes the pairwise elasticity between s_L and s_H .

Proof. See Appendix.

To simulate the optimal policy empirically we need three types of objects: social welfare weights g_i , elasticities, and the joint distribution of income, medical spending, and health. In section 4, we estimate the latter two objects, and in section 4, we present the optimal policy results under various specifications of the social welfare weights.

3 Behavioral Responses and the Gradient of Health and Wealth

Disparities in health and healthcare across socioeconomic status are a well documented phenomena in the literature, but less is known about how *behavioral responses* to health insurance coverage may differ among the rich and the poor. We investigate this question using experimental data from the RAND Health Insurance experiment, testing for heterogeneity in the elasticity of medical spending across socioeconomic status by separating individuals into income quintiles. Consistent with what theory might predict, we find that individuals in the bottom quintiles do not really respond to cost-sharing incentives, while individuals in the middle and top income quintiles respond substantially.

Intuitively, individuals at the bottom of the income distribution are living in poverty and do not have the time or energy to over-consume medical services, or to respond to marginal changes in the generosity of health insurance. These individuals living at or below the Federal Poverty Line likely go to the doctor when it is necessary, and their medical bill often ends up higher than what they can afford to pay. This is why government programs such as Disproportionate Share Hospital (DSH) payments and the 340B drug discount program have stepped in to compensate, monetarily, the healthcare providers that care for the poor.

Our empirical exercise involves taking one large experiment and cutting it up into five experiments, where the first order concern may be that of power. Luckily, random assignment and balance across plans along income covariates provide us with enough statistical power to reject the hypothesis that the elasticity of medical spending is the same across income quintiles. We have not seen other papers do this particular decomposition of the RAND elasticity by socioeconomic status, and, despite its simplicity, we believe it is informative and of first order importance to health insurance policy design, particularly for governments that care about inequality and socioeconomic disparities in healthcare. That said, we view our marginal contribution as largely instrumental to our main objective of the paper by allowing us to account for behavioral responses to health insurance policies in our characterization of the optimal policy.

Data Description: We use data from the RAND Health Insurance experiment, which was a large scale field experiment conducted in the 1970's designed to estimate and quantify moral hazard. The experiment randomly assigned individuals to health insurance plans with different levels of generosity, and aimed to establish the causal effect of plan generosity on total healthcare spending. The data have been originally published in the public domain by Newhouse and his team of researchers, and carefully revisited in Aron-Dine et al. (2013). We use the data processed by Aron-Dine et al. (2013), and borrow generously from their code, in order to estimate the elasticities of medical spending to coinsurance by socioeconomic status.

The RAND experiment ran from 1974 to 1981, and enrolled over 5,800 individuals. The experiment randomly assigned individuals to one of six plans, which varied the level of cost-sharing or deductible. The first four plans differed in that they set overall coinsurance rates at 95%, 50%, 25%, or 0%; the fifth plan had a mixed coinsurance rate of 50% for dental and mental health services and 25% for everything else; and the sixth plan, referred to as the "individual deductible" plan, had 95% coinsurance for outpatient services, and 0% for inpatient. The experiment ran in six different locations, and random assignment was random conditional on location and starting month. The experiment collected data on individual's income prior to enrollment, and excluded from the experiment individuals with incomes greater than \$25,000 in 1973 dollars (which corresponds to

approximately \$167,000 in 2022 dollars).

Decomposing the RAND Elasticities by Socioeconomic Status

Empirical Framework: In our main specification, we estimate elasticities by running the standard regression of log-spending on log-price, interacted with income quintile. We follow the RAND investigators and Aaron-Dine et al. (2013) and use person-year as the primary unit of analysis. We denote individuals by i, calendar year by y, medical spending by m, the insurance plan generosity by s, insurance plan by p, income by z, income quintile by q, month by t, and location by l.

$$\log(m_{i,y}) = \eta_q(\log(s_p) \times \mathbb{1}[z_i \in q]) + \beta_y + \beta_{l,t} + \epsilon_{i,y}$$

Insurance plan generosity only takes one of four values in the RAND experiment. Location by month fixed effects are included because random assignment was only random conditional on these factors. The standard, average elasticity of medical spending with respect to insurance plan generosity will correspond to the estimated η_q coefficients. Figure 3 shows the results from this regression. It appears evident that the behavioral response to insurance incentives is higher among individuals with more income.



Figure 3: Usual elasticities by Income Quintile

Note: Elasticities are calculated based on regressions of log(medical spending + 1) on log(price), where price is either the coinsurance rate of the plan or the average out of pocket price paid by individuals assigned to that plan.

	Income Quintile				
	Bottom	Second	Third	Fourth	
	p-values				
Second	0.103				
Third	0.000	0.041			
Fourth	0.000	0.000	0.0213		
Тор	0.000	0.000	0.000	0.0281	
Joint F-Test	18.89	13.52	9.56	4.83	
p-value	0.000	0.000	0.0001	0.0281	

Table 2: Pairwise Parameter Tests for Equality in the Elasticity Estimates

We then conduct a joint F-test and reject the null that all five coefficients are equal. We also conduct a Wald test on all the pairwise comparisons of parameter estimates, shown in Table 3. While we cannot reject that the elasticity is the same for individuals in the bottom two income quintiles, the evidence from the RAND experiment seems to support the hypothesis that behavioral responses to insurance incentives differ along the income distribution.

Threats to Validity: The literature following the original RAND experiment has raised concerns about potential threat to the validity of the estimates, and we discuss these concerns as they pertain to our empirical exercise. In particular, researchers have investigated whether the stratified random assignement was indeed random (Aaron-Dine et al., 2013), and whether attrition may have biased the estimates (Newhouse, 1993; Nyman, 2008; Aaron-Dine et al., 2013; Newhouse et al., 2008). The conclusions from these studies are that assignment of plans to individuals was indeed random, with the possible exception of the 50% coinsurance plan, which had relatively few people assigned to it and was discontinued midway through the experiment. Importantly for us, there are no statistically significant correlations between plan characteristics and individual or household income (Manning et al., 1987; Aaron-Dine et al., 2013), and this is because the assignment algorithm was designed to achieve balance on income, among other characteristics, across plan assignment.

The conclusions from studies on potential attrition bias in the RAND estimates, however, are somewhat different. Attrition, or differential participation across individuals, could bias estimates in one of two ways. The estimates could be biased upward if it was the case that individuals with high ex-ante expected spending only agreed to participate if assigned to a generous plan, while individuals with low ex-ante expected spending agreed to participate, regardless of plan assignment. Conversely, estimates could be biased downward if there were unobservable fixed costs to participate if and the experiment: individuals with high ex-ante expected spending agreed to participate agreed to participate fixed costs to participate.

regardless, whereas individuals with low ex-ante expected spending may have only participated when assigned to a generous plan.

The original RAND investigators address potential attrition bias by examining refusal of enrollment and withdrawal from the experiment, and they conclude that neither of these forces caused any appreciable bias (Newhouse, Chapter 2; 1993). Aaron-Dine et al. (2013) investigate attrition by implementing an omnibus test for differences in observable pre-randomization characteristics among individuals who completed the experiment. They find statistically significant imbalances across plans on two characteristics: number of doctor's visits prior to the experiment and share of participants who had a medical exam in the year prior to the experiment. In light of attrition concerns, they take a bounding approach, conduct a statistical exercise to estimate a lower bound for the treatment effect following Lee (2009). They conclude that the RAND data still reject the null hypothesis of no healthcare utilization response to cost sharing, but caution that the magnitude of the estimates could be considerably lower than the unadjusted estimates.

For the purposes of our present exercise, we should be concerned about attrition bias if it were the case that the underlying economic mechanisms driving this bias differed across income quintile. For instance, we may be concerned that individuals from the bottom income quintile had differential participation rates into plan due to fixed costs of participation, while individuals in the top income quintile had differential participation due to ex-ante expectations about medical spending. In this case, differential attrition would would bias downward the elasticity estimated on for the bottom quintile, and upward the elasticity estimate for the top quintile. The bounding exercise in Aaron-Dine et al. (2013) remains agnostic about the underlying economic mechanisms, and we can similarly estimate bounds on the treatment effect separately, by income quintile. Given limitations on sample size, we are underpowered on this bounding exercise. Nonetheless, we can still reject the null that the lower bound of the elasticity is the same across all income quintiles.

Elasticities at Various Points of the Price (Coinsurance) Schedule

When we simulate the optimal policy, we use pairwise elasticities at various points of the coinsurance schedule, as opposed to average elasticity. This is because we wish to capture meaningful differences in the elasticity along various points of the price (coinsurance) schedule. Given that our primary objective is to characterize optimal insurance policy, and we characterizing the optimum by looking at local deviations around a particular policy, our simulations use the pairwise elasticity estimates that most directly speak to the local deviations of the policy being evaluated.

For example, consider a local perturbation to Medicaid's 100% health care subsidy so that insurance is now less generous (e.g. 80%). In order to estimate the marginal fiscal externality from making insurance less generous (and the government's savings from cutting back Medicaid), we would like to account for the possibility that individuals may respond discontinuously to 20%

coinsurance versus free care, relative to a policy reform that changes the coinsurance rate from 50% to 20%. We would also like to account for the fact that an individual in the bottom income quintile will not consume significantly less care under the less generous policy, relative to the previous policy, so the magnitude of the counterfactual savings may not be as large as one would estimate with the average elasticity. In general, the magnitude of the fiscal externality will depend on the incidence of the policy reform. Given that the RAND experiment allows us to account for discontinuous responses along the (out-of-pocket) price schedule, and for heterogeneous responses along the income distribution, we incorporate this information into our simulations.

Given that we rely entirely on these pair-wise arc elasticities, we discuss our estimation procedure in detail here. Pairwise elasticities are calculated as percent change in total spending, divided by the percent change in coinsurance.

$$\eta(s,s') = \frac{\mathbb{E}[m_i|s'] - \mathbb{E}[m_i|s]}{\mathbb{E}[m_i|s]} \cdot \frac{s}{s'-s}$$

Pairwise elasticity standard errors are bootstrapped based on 500 replications and clustered on family. That is, the $\mathbb{E}[m_i|s]$ are estimated based on regressing m_i on plan dummies λ_p and location by month fixed effects, clustering standard errors at the family level. We compute pairwise elasticities instead of arcelasticities, and we partition the sample into five based on the individual's income quintile. We compute the pairwise elasticities by comparing the 0%, 25%, 50%, 95% coinsurance plans. The estimates from this procedure can be seen in Figure 3.



Figure 4: Pairwise Elasticities by Income Quintile

First, consistent with Aaron Dine et. al. (2013), we find that the elasticity is discontinuous at the free care plan. If we decompose the behavioral response to insurance around the free care plan by income, we paint a slightly different story than what the log-log elasticity estimates suggested:

	Bottom Quintile	Top Four Quintiles		
25vFC	-0.552	-1.330***		
	(0.828)	(0.275)		
50vFC	0.0413	-0.643***		
	(0.849)	(0.169)		
95vFC	-0.226	-0.422***		
	(0.188)	(0.0511)		
25v50	-0.184	-0.0166		
	(0.811)	(0.154)		
25v95	-0.0320	-0.0366		
	(0.128)	(0.0381)		
50v95	-0.257	-0.130		
	(0.670)	(0.131)		
N	3813	16390		

Standard errors in parentheses

* p < 0.05, ** p < 0.01, *** p < 0.001

Table 3: Pairwise Elastic	tities by Income Cutoff
---------------------------	-------------------------

low income individuals are not as responsive to coinsurance, but *neither are very rich individuals*. This likely captures the fact that wealthy individuals have high enough levels of income already, so the marginal health care subsidy only has a second order effect on their medical spending decisions.

Second, the 50% coinsurance plan was discontinued throughout the experiment and had a substantially smaller number of participants than other plans, which explains why the standard errors are wide. However, the pairwise elasticity estimate that comes from comparing the 25% to the 95% coinsurance plan is a much tighter zero, suggesting that the behavioral response along medium to high levels of coinsurance may not be that sensitive to the specific level of coinsurance. In our baseline optimal policy simulations, we use the the point estimates from the pairwise comparisons that include the 50% plan and using these in our calibrations. However, our results are robust to using the point estimates around the 25% to the 95% interval, instead.

In practice, we end up grouping individuals based on whether they fall above or below a given income quintile. The reason we do so is that the our optimal policy depends only on the aggregate elasticity above and below the income threshold that qualifies individuals for a health insurance safety net, such as Medicaid. In the US, public health insurance provided by Medicare and Medicaid conditions their insurance policy on whether individuals fall above or below 133% of the Federal Poverty Line, which roughly corresponds to the bottom income quintile.

Table 3 shows the pairwise elasticity estimates across these two groups. The story is the same:

individuals in the top four income quintiles are, on average, more responsive to health care subsidies than individuals in the bottom quintile. Behavioral differences along the socioeconomic gradient provide us with the key empirical objects that we need in order to quantify the fiscal externality (i.e. the 'budget effect') from a particular health insurance policy reforms like making insurance more generous or changing the Medicaid eligibility threshold. In the next section we characterize the joint distribution of medical spending, income, and health.

What Income and Medical Spending Tell Us About Health

There is a vast literature in health economics and macroeconomics that has documented health and healthcare disparities across socieconomic status, and what we show here is both consistent and by no means novel or surprising. Nonetheless, the empirical relationship between income, medical spending, and health is relevant for our optimal policy. We take the underlying distribution of health states as exogenous, and hold the relationship fixed when characterizing the counterfactual optimal policy.⁴

We examine the empirical relationship between income, consumption, and health using data from the Medical Expenditure Panel Survey (MEPS). Our objectives in this section are twofold: first, we want to showcase patterns in our data that capture the relationship between health disparities and the two "taxable instruments" that our social planner can influence– income and health. Second, we want to estimate the objects that directly go into our policy formulas–the covariance of health and medical spending for high and low income individuals, and the health and income.

Data Description: The MEPS data contain information about the individual's income, health insurance, aggregate medical expenditures (aggregate and decomposed by inpatient, and outpatient, out-of-pocket), age, illnesses, and various survey responses that attempt to capture the individual's overall health rating. Individuals in the MEPS survey report their overall health state on a scale from one to five, and they also fill out a 36 question survey (the Short Form-26) which produces two composite scores that summarize the individual's mental health and physical health. These are the two measures that we primarily use to estimate H. The MEPS data consist of a large-scale survey of individuals and families, and it is panel data.

The MEPS data are made publicly available for download by the Agency for Healthcare Research and Quality. We use data from 2009 through 2016 because the data contain more detailed information about health medical spending (data after 2016 do not provide BMI, and data prior

⁴An important caveat to our paper is that we take the health distribution as exogenous. In practice, however, Health outcomes may be largely determined by factors outside our model, such as the private health insurance market, access to health care (Fogel and Lee, 2003), differences in quality (Skinner and Zhou, 2004), or even differential physician behavior toward the rich and the poor (Chen and Lakdawalla, 2019). Explicit consideration of these factors is somewhat beyond our model. Nonetheless, our thought experiment of the optimal health care subsidy does not preclude that these other factors co-exist. Rather, we need to assume that public health care subsidies do not directly impact these outside factors so as to alter the underlying distribution of health.

to 2009 do not separately break out inpatient and outpatient spending, for example). We use two types of data files: the Full Consolidated file and the Medical Conditions File. From the Full Consolidated file, we obtain income, self reported health rating scores, aggregate medical spending, and consumption (where we define consumption as income net of out of pocket medical spending). From the medical conditions file, we obtain the diagnosis codes (ICD-9) associated with each individuals. The medical conditions file allows us to construct indicators for the 17 disease categories that factor into the Charlson Commobidy Index, which is a commonly used index in the health literature, used to characterize the severity of illness of an individual. We borrow code from Frakes and Gruber (2019) to calculate Charlson scores from MEPS data.

Table 3 contains standard summary statistics describing the individuals in our sample. Relative to the top income quintile, low income individuals in our sample are younger and low income women are over-represented.

The MEPS survey design includes stratification, clustering, multiple stages of selection, and disproportionate sampling. The MEPS data provider creates weights that account for the non-random sampling, allowing us to correct the point estimates so as to obtain nationally representative aggregates. The means reported in the Table below have been adjusted appropriately based on the survey design, weights, and strata.

Measuring Quality of Life

In the Grossman model of health capital, quality of life (QoL) describes a scalar between zero and one that scales down the flow payoff utility of an individual based on whether they suffer of illnesses or chronic conditions. There is a literature in health economics that has devoted explicit attention to empirically estimating Quality Adjusted Life Expectancy (QALE) and Quality Adjusted Life Years (QALYs), which map the objects in the theory– the H_i scalar and the survival probabilities– to data. As noted in Cutler et. al. (2022), there is no universally accepted metric of health-related QoL, and it is usually measured in a series of "domains," including physical and mental functioning, role limitations, pain, and cognition.

One common approach to estimating H_i involves regressing survey measures of self-reported health on a vector of health conditions. Cutler and Richardson (1999), for instance, use this approach to compare the effects of various health conditions (e.g. arthritis, diabetes, hypertension) on Quality of Life. In a more recent paper, Cutler et. al. (2022) combine data on survival probabilities, self-reported health, and a number of other sources to estimate QALE at age 65. Within this literature, the focus is usually on comparisons across disease categories and particular medical interventions. QALYs are used by the United Kingdom's National Health Service when deciding which pharmaceutical treatments to pay for. Our exercise follows the same methods.

Empirical Framework: In our baseline specification, we estimate QoL, H, by regressing the overall,

	Income Ouintile					
	Bottom	Second	Third	Fourth	Тор	
Sociodemographics						
Age	41.53	48.96	47.86	48.11	49.93	
Male	0.374	0.439	0.477	0.524	0.599	
White	0.738	0.785	0.814	0.831	0.838	
Black	0.160	0.141	0.117	0.0996	0.0711	
Other	0.101	0.0739	0.0686	0.0695	0.0905	
Married	0.451	0.425	0.532	0.614	0.679	
Education (years)	8.387	8.521	9.267	10.01	11.02	
Behavioral Risk Factors (share	es)					
Smoker	0.198	0.194	0.173	0.138	0.0855	
BMI Normal	0.378	0.339	0.318	0.313	0.331	
BMI Overweight	0.275	0.315	0.332	0.342	0.371	
BMI Obese	0.278	0.287	0.301	0.304	0.265	
Comorbidity Factors						
Asthma	0.136	0.118	0.0969	0.0925	0.0888	
Cancer	0.0979	0.148	0.126	0.118	0.127	
Diabetes	0.125	0.144	0.116	0.0975	0.0745	
Heart Attack	0.0488	0.0678	0.0488	0.0332	0.0268	
Charlson Score	0.326	0.391	0.287	0.237	0.226	
Health Status (nercentage points relative to the mean)						
Self-reported, Overall	-1.81	-3.81	0.36	2.90	7.20	
Mental Health	-2.31	-1.21	0.89	1.80	2.81	
Physical Health	-1.97	-3.25	0.42	2.65	4.13	
Components of Quality of Life						
Hearing Impairment	0.0579	0.0887	0.0728	0.0640	0.0586	
Writing Impairment	0.0375	0.0007	0.0720	0.0040	0.0228	
Health Limits Social Activity	0.0005	0.107	0.0557	0.0352	0.0220	
Difficulty Lifting 10 lbs	0.117	0.150	0.0724	0.0402	0.0303	
Difficulty Crouching	0.143	0.172	0.0757	0.0354	0.0372	
Difficulty Reaching	0.133	0.162	0.0933	0.0550	0.0385	
· · · · · · · · · · · · · · · · · · ·						
Ν	41,409	38,654	31,425	27,030	22,747	

Table 4: Summary Statistics by Income Quintile

1-5, self reported health rating on a vector of conditions and obtaining the predicted values, where the outcome variable is rescaled to lie between zero and one. The unit of analysis for all of our empirical estimates is person-year. Denoting individuals by i, years by t, age by a_t we estimate:

$$H_{it}^{self-reported} = \beta_{cc} + bmi_{it} + \alpha \cdot X_{it} + \beta_t + \epsilon_{it}$$
(4)

where cc denotes a particular chronic condition, bmi_{it} denotes the individual's body mass index, X_{it} denotes a vector of demographic variables which include gender, age, age squared, age interactions with gender. The chronic conditions vector includes cancer, high blood pressure, coronary artery disease, angina, an indicator for whether the individual has experienced a heart attack, an indicator for a prior stroke, history of heart disease, emphysema, high cholesterol, diabetes, arthritis, asthma, and the various functional impairments (i.e. difficulty walking, reaching, bending over, writing, hearing, attending social activities).

Figure 3 shows the predicted values from the regression in equation (4) against individual medical spending, age, and sickness metric (Charlson Commorbidity Index). The upper left panel shows QoL by income decile conditional on various categories of medical spending. Conditional on having zero health care spending, QoL seems to be fairly constant across income deciles, suggesting that differential levels of consumption are the primary source of disparities among these individuals. Conditional on any type of spending, however, a clear pattern of health disparities emerges. Individuals in the third income decile are, on average, worse off than everyone else. One possible explanation for this dip is that individuals in the third income quintile are not poor enough to qualify for social assistance programs or Medicaid, but they still have low levels of income and consumption.



Note: The lower right panel only contains data from 2009 through 2015 due to data availability in the MEPS. The remaining three panels contain data from 2009 through 2016.

Figure 5: Predicted Health Status (Self-Reported Overall)

The upper right panel of Figure 3 shows QoL by *medical spending* decile, conditional on being in the top versus the bottom quintile. As expected, high income individuals have better QoL, even among the high medical spending individuals (who are presumably suffering of more illness than low spending individuals). The bottom left panel shows QoL by age across the top and bottom income quintiles. The gap in QoL disparities is highest for individuals in their 40's and 50's, and the gap shrinks somewhat after the retirement age. Finally, the lower right panel shows disparities in QoL for individuals with positive Charlson scores, which capture the number of commodities affecting an individual. Conditional on having no comorbidities, the income disparities in QoL are smaller than among individuals with positive comorbidities. This last panel is suggestive of a rationale behind a public health insurance safety net that caters to the chronically ill, such the public health care system in Italy, Japan, and the Netherlands.

4 The Optimal Healthcare Subsidy and Safety Net Eligibility

Our optimal policy requires that we specify three types of objects: elasticities, the joint distribution of health, income, and medical spending, and the social welfare weights. We have already described how we obtain the first two objects, which leaves the social welfare weights are the final unknown. The social welfare weights capture social preferences, and they come from social, moral, or philosophical concepts outside the scope of our model. In the traditional welfare economics literature, the two most common specifications for social preferences have been utilitarianism (everyone's utility is weighted equally) and Rawlsian (the individual who is worse off in society gets all the weight). These two objectives, however, lead to optimal policies that seem far removed from policies used by governments across the globe, in practice. Saez and Stantcheva (2016) provide an extensive discussion of different types of social preferences that more closely align with the policies seen in practice. In the context of taxation, they discuss social objectives such as poverty alleviation, equality of opportunity, Libertarianism, and the implications that these various objective have for tax policy. Our exercise builds off of the foundations laid out in their work, and we begin with a social objective of health equity.

In this section, we simulate the optimal health care subsidy and tax policy for a planner who cares about *health equity*. By health equity, we refer to reducing disparities that exist in health-capital-adjusted utility, taking the underlying distribution of health as a fixed primitive of the population. Our social planner will thus choose a policy based on an objective that places a higher weight on sick individuals relative to healthy individuals, trading off the costs of subsidizing health care against the fiscal externalities of financing such subsidies.

Estimation Algorithm

To simulate the optimal policy, first we specify the social welfare weights as a function of observable characteristics. For our main specifications, we use weights that are inversely proportional to only health, or Rawlsian in to health adjusted income.

$$g_i = -\log(H_i)$$
 or $g_i = \min\{H_{it} \cdot z_i\}$

Our unit of analysis is person-year. We take the actual self reported health score for individual i in year t that was used for estimating equation (4).⁵ That is, we use H_{it} , which has been normalized to lie between zero and one. We inflation-adjust income, medical spending, and consumption (income net of out of pocket spending) to 2010 dollars.

The optimal tax, τ , depends on \bar{g}_z and the labor supply elasticity ξ_z . The \bar{g}_z is equal to product of the expectation of the social welfare weight and income, divided by the average income and

 $^{^{5}}$ We use the actual values, as opposed to the predicted values, to reduce possible sources of error in the estimation.

welfare weights, $\bar{g}_z = \mathbb{E}[g_i \cdot z_i]/(\mathbb{E}[z_i] \cdot \mathbb{E}[g_i])$. We compute $\mathbb{E}[g_i \cdot z_i]$, $\mathbb{E}[z_i]$, and $\mathbb{E}[g_i]$ at the specified weights using the survey data estimation software in Stata, accounting for the sampling weights and strata given by the data provider. We calibrate the labor elasticity at a constant 0.5, which is on the upper end of the range of estimates in the literature, following the calibrations done in Saez and Stantcheva (2016).

The optimal subsidy for low and high income individuals depend on \bar{g}_m^L , \bar{g}_m^H , and the elasticities of medical spending with respect to coinsurance, conditional on income, $\eta_m(s|z_i \leq \hat{z})$ and $\eta_m(s|z_i > \hat{z})$. The \bar{g}_m^L is similar to the \bar{g}_z but with respect to m_i , but the main difference is that it is the conditional expectation for low income individuals, $\bar{g}_m^L = \mathbb{E}[g_i \cdot m_i | z_i \leq \hat{z}]/(\mathbb{E}[z_i | z_i \leq \hat{z}] \cdot \mathbb{E}[g_i])$. The \bar{g}_m^H is almost identical to \bar{g}_m^L , but the conditional expectation is over the high income individuals, $\bar{g}_m^H = \mathbb{E}[g_i \cdot m_i | z_i > \hat{z}]/(\mathbb{E}[z_i | z_i > \hat{z}] \cdot \mathbb{E}[g_i])$. Notice that the expectation of the social welfare weights in the denominator is *unconditional*; this is because it comes from the multiplier on the government budget constraint λ , which is pinned down by the lump sum transfer which applies to everyone. The empirical estimate of the \bar{g}_m 's thus will differ depending on the level of \hat{z} . We estimate the \bar{g}_m 's at every percentile of z_i in order to search over the space of \hat{z} that yield the optimal policy. We compute the optimal s_L and s_H at every possible \hat{z} , and we discretize the income space into 100 quantiles for this computation.

The s_L and s_H depend on the also on the elasticities, $\eta_m(s|z_i \leq \hat{z})$ and $\eta_m(s|z_i > \hat{z})$, which we take from RAND. We calculate pairwise elaticities between the health care subsidies of 100%, 75%, 50% and 5% (i.e. the coinsurance rates 0%, 25%, 50%, and 95%) by income subgroup. Due to power issues, we discretize income into quintiles for this computation, and estimate aggregate pair-wise elasticities for individuals below \hat{z} by averaging over individuals in the bottom quintile, the bottom two quintiles, the bottom three quintiles, the bottom four quintiles, and for the full sample. We estimate aggregate pair-wise elasticities for individuals above \hat{z} in the same fashion. For each computation of s_L and s_H given \hat{z} , we use the empirical elasticity estimates for quintile of that particular \hat{z} (e.g. given a \hat{z} in the 15th percentile, we would use the aggregate elasticity estimated off of the bottom RAND income quintile for $\eta_m(s|z_i \leq \hat{z})$, and the aggregate elasticity estimated off of the top four RAND income quintiles for $\eta_m(s|z_i > \hat{z})$). We then compute the gradient of the Lagrangian using the elasticity estimates at the two corners to check whether an interior solution for s exists.⁶ If this derivative is positive at the upper bound of the subsidy, we say s = 1, and if it is negative at the lower bound of the subsidy, we say s = 0. If an interior solution does exist, we compute s_L and s_H for the three local pairwise elasticities (i.e. 100%, versus 75%, 75% versus 50%, and 50% versus 5%), and then check whether the empirical values of s_L and s_H at each of the three elasticities are internally consistent with the range in which the pairwise elasticity was estimated (e.g. if the s_L estimated using the pariwise elasticity between 100% and 75% health care subsidy is equal to 60%, we reject that estimate of s_L). When no values of s_L and s_H are internally consistent, we take the estimate that is closest to the boundary point of the

⁶The derivative of the Lagrangian with respect to s is equal to $\bar{g}_m^L - (\eta_m(s|z_i \leq \hat{z}) + 1)$.

pairwise range.

Finally, we search over the space of \hat{z} to recover an optimal policy. The optimal \hat{z} depends on the average marginal social welfare weight of individuals at the threshold, $\bar{g}_m(\hat{z})$, and the medical spending elasticity with respect to coinsurance for those individuals, $\eta_m(s_L, s_L - s_H | \hat{z})$. The $\bar{g}_m(\hat{z})$ is again very similar to the other welfare weights, with the only difference that it conditions over a different range. Since we have discretized the income space into 100 quantiles, we take the conditional expectations among individuals within each quantile. That is, we estimate $\bar{q}_m(\hat{z}) =$ $\mathbb{E}[g_i \cdot m_i | z_i \in pctile(\hat{z})]/(\mathbb{E}[m_i | z_i \in pctile(\hat{z})] \cdot \mathbb{E}[g_i])$, where $pctile(\hat{z})$ denotes the range of incomes that lie within a given percentile. For the elasticity, $\eta_m(s_L, s_L - s_H | \hat{z})$, we estimate all possible pairwise elasticities in the RAND data for each income quintile and construct a four by four matrix, where the rows and columns index subsidies of 100%, 75%, 50%, or 25%. We then find the pairwise comparison in the RAND experiment that is closest to the estimated values of s_L and s_H using estimates from the income quintile where the conjectured \hat{z} lies. Next, we compute the optimality condition for \hat{z} at the corners in order to check whether there is an interior solution. If the optimality condition is negative at first income percentile, we say the optimal $\hat{z} = 0$, and if it's positive at the 100th income percentile, we say $\hat{z} = Z$ (Z is the upper bound of the income distribution). If there is an interior solution, we take the minimum over the absolute values of the optimality conditions across the 100 possible values of \hat{z} .

Finally, we take the optimal \hat{z} , the optimal s_L and s_H for that particular \hat{z} , and the optimal linear tax τ , to calculate the transfer, R. Given the binding budget constraint, $R = \tau \hat{z} - s_L \mathbb{E}[m_i|z_i \leq \hat{z}] - s_H \mathbb{E}[m_i|z_i > \hat{z}]$. To calculate the level of R, we use the average income and conditional averages of medical spending above and below the optimal \hat{z} .

Results

Table 4 below shows our results. The the welfare weights in columns (1) and (4) are inversely proportional to health-adjusted income. The $\log(H_i)$ weights in column (1) result in a policy that is strikingly similar to Medicare and Medicaid– Medicare covers 80% of expenses for outpatient care, and 100% for low income individuals that qualify for Medicaid. While Medicaid eligibility varies by state, the US Congress incentivizes states to choose 133% of the Federal Poverty Line or above as the Medicaid eligibility threshold by disbursing federal subsidies to these states. The Medicare deductible for hospital care is around \$1,500, which is fairly similar to the lump sum tax suggested by the policy in column (1). The $\frac{1}{H_i^2}$ weights in column (4) describe a planner who places even *more* weight on the sick, relative to the health (i.e. the "concavity parameter" with respect to health is greater). Under these weights, the planner optimally chooses "Medicare for All except for the top 1%"– subsidize 100% of care for all individuals below 16,000% of the Federal Poverty Line, which corresponds to the 99th income percentile. Because low income individuals also have poorer health, the optimal policy in column (4) also involves a transfer of \$4,400, which implies

	Social Welfare Weight Specification g_i					
	(1)	(2)	(3)	(4)	(5)	
	$-\log(H_i)$	$\min\{H_i \cdot z_i\}$	$\frac{1}{H_i \cdot z_i}$	$\frac{1}{H_i^2}$	$\min\{H_i\}$	
High Inc Subsidy, s_H	0.704	0.378	0	0	1	
Low Inc Subsidy, s_L	1	1	0	0.785	1	
Eligibility as % FPL, \hat{z} (Inc Percentile)	129% (11th pctile)	309.7% (45th pctile)		16,020% (99th pctile)	•	
Tax, τ	0.198	0.452	0.667	0.235	0.52	
Demogrant, R	-\$1,959.4	\$ 9,054.6	\$ 23,692	\$ 4,404	\$7,662.4	

that all individuals who earn less than \$18,700 a year receive net transfers from the government.

Table 5: Simulated Optimal Policy

Columns (2) and (5) show the Rawlsian welfare weights. The planner in column (2) cares about the poorest individuals with the worst quality of life, while the planner in column (5) cares only about the individuals in society who have the lowest quality of life. The Rawlsian planner in column (2) chooses a Medicaid-type policy, but one that is substantially more generous than the US, with the eligibility threshold at 309% of the Federal Poverty Line (which corresponds to approximately the 45th income percentile). The Rawlsian planner in column (5) chooses "Medicare for All" and gives a relatively high transfer to everyone of \$ 7,600. This means that individuals who make less than \$14,000– which is close to the Federal Poverty Line for a one person household individual–receive net transfers from the government.

Column (3) shows welfare weights that are inversely proportional to income. Under these weights, the planner optimally chooses a large transfer and tax rate– 67% marginal tax rate and nearly a \$24 K transfer– which means that individuals earning less than \$35,000 a year receive net transfers from the government, and individuals above that figure are taxed.

5 Conclusion

Even the wealthiest individuals in a nation would find themselves incapable of enjoying their wealth if their health condition were one of severe illness. Socially efficient insurance is typically characterized as satisfying a trade-off between the value of risk protection and the cost of induced moral hazard, but we provide a new lens under which to evaluate public health insurance– redistribution. The standard insurance-moral hazard trade-off is based on maximizing unweighted social surplus. However, when there is heterogeneity among the recipient population, either in existing levels of health, or along some other dimension such as income, insurance also involves direct cross-type redistribution in addition to within-individual redistribution across realized (e.g. health) states. We view are contribution as providing a new toolkit to the design and evaluation of public health insurance.

An important caveat to our exercise is that we have completely abstracted away from the *in-surance* aspect of public health insurance. In our model, individuals know their type and make optimal choices based on this information, while in reality there are salient information frictions in health care that make insurance valuable in the market. Here, we have ignored the existence of a private insurance market, which begs the question of why have government intervention in the first place. To this point, we would say that our exercise is not about addressing market failures in the health care industry, but rather about how to design a tax and transfer system for a planner who has concerns for health equity, and in the same way that the Cadillac tax acts through the "tagging" channel, medical spending serves as a tag for sick individuals. We do recognize, however, that it is important to account for the interplay between public health insurance and the private insurance market, and leave this open for future work.

References

- Aron-Dine, Aviva, Liran Einav, and Amy Finkelstein, "The RAND Health Insurance Experiment, Three Decades Later," *Journal of Economic Perspectives*, February 2013, 27 (1), 197–222.
- **Arrow, Kenneth**, "Welfare Analysis of Changes in Health Coinsurance Rates," in "The Role of Health Insurance in the Health Services Sector," NBER, 1976, pp. 1–34.
- **Bajari, Patrick, Christina Dalton, Han Hong, and Ahmed Khwaja**, "Moral hazard, adverse selection, and health expenditures: A semiparametric analysis," *The RAND Journal of Economics*, 2014, 45 (4), 747–763. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1756-2171.12069.
- Barnichon, Regis and Yanos Zylberberg, "A Menu of Insurance Contracts for the Unemployed," *The Review of Economic Studies*, May 2021, p. rdab026.
- Bierbrauer, Felix J., Pierre C. Boyer, and Andreas Peichl, "Politically Feasible Reforms of Nonlinear Tax Systems," *American Economic Review*, January 2021, *111* (1), 153–191.
- Blomqvist, Ake, "Optimal non-linear health insurance," *Journal of Health Economics*, June 1997, *16* (3), 303–321.
- Blundell, Richard, Mike Brewer, Peter Haan, and Andrew Shephard, "Optimal Income Taxation of Lone Mothers: An Empirical Comparison of the UK and Germany*," *The Economic Journal*, 2009, *119* (535), F101–F121. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0297.2008.02261.x.
- **Bundorf, M. Kate and Daniel P. Kessler**, "The Responsiveness of Medicaid Spending to the Federal Subsidy," November 2021.
- **Cardon, James H. and Igal Hendel**, "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey," *The RAND Journal of Economics*, 2001, *32* (3), 408–427. Publisher: [RAND Corporation, Wiley].
- **Chandra, Amitabh, Evan Flack, and Ziad Obermeyer**, "The Health Costs of Cost-Sharing," Working Paper 28439, National Bureau of Economic Research February 2021. Series: Working Paper Series.
- **Chang, Tom**, "What do Nonprot Hospitals Maximize? Evidence from Californias Seismic Retrot Mandate," p. 64.
- Chang, Yongsung and Yena Park, "Optimal Taxation with Private Insurance," *The Review of Economic Studies*, November 2021, 88 (6), 2766–2798.
- **Chen, Alice and Darius N. Lakdawalla**, "Healing the poor: The influence of patient socioeconomic status on physician supply responses," *Journal of Health Economics*, 2019, *64* (C), 43–54. Publisher: Elsevier.
- Chetty, Raj and Emmanuel Saez, "Optimal Taxation and Social Insurance with Endogenous Private Insurance," *American Economic Journal: Economic Policy*, May 2010, 2 (2), 85–116.
- ____, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler, "The Association Between Income and Life Expectancy in the United States, 2001-2014," JAMA, April 2016, 315 (16), 1750–1766.

Cutler, David, "Health care and the public sector," Handbook of Public Economics, Elsevier 2002.

- _ and Elizabeth Richardson, "Your Money and Your Life: The Value of Health and What Affects It," *Frontiers in Health Policy Research*, February 1999.
- and _, "Your Money and Your Life: The Value of Health and What Affects It," Frontiers in Health Policy Research, February 1999.
- **Deaton, Angus**, "Health, Inequality, and Economic Development," *Journal of Economic Literature*, March 2003, *41* (1), 113–158.
- _, The Great Escape September 2013.
- **Deaton, Angus S. and Christina Paxson**, "Health, Income, and Inequality over the Life Cycle," in "Frontiers in the Economics of Aging," University of Chicago Press, January 1998, pp. 431–462.
- **Dunn, Abe, Joshua D. Gottlieb, Adam Shapiro, Daniel J. Sonnenstuhl, and Pietro Tebaldi**, "A Denial a Day Keeps the Doctor Away," July 2021.
- **Dworczak, Piotr, Scott Duke Kominers, and Mohammad Akbarpour**, "Redistribution Through Markets," *Econometrica*, 2021, 89 (4), 1665–1698. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA16671.
- Ellis, Randall P. and Willard G. Manning, "Optimal health insurance for prevention and treatment," *Journal of Health Economics*, December 2007, *26* (6), 1128–1150.
- Feldstein, Martin and Bernard Friedman, "The Effect of National Health Insurance on the Price and Quantity of Medical Care," in "The Role of Health Insurance in the Health Services Sector," NBER, 1976, pp. 505–541.
- and Jonathan Gruber, "A Major Risk Approach to Health Insurance Reform," in "Tax Policy and the Economy, Volume 9," MIT Press, January 1995, pp. 103–130.
- **Finkelstein, Amy, Erzo F. P. Luttmer, and Matthew J. Notowidigdo**, "WHAT GOOD IS WEALTH WITHOUT HEALTH? THE EFFECT OF HEALTH ON THE MARGINAL UTIL-ITY OF CONSUMPTION," *Journal of the European Economic Association*, January 2013, *11*, 221–258.
- _, _, and _, "What Good Is Wealth Without Health? The Effect of Health on the Marginal Utility of Consumption," *Journal of the European Economic Association*, 2013, *11* (s1), 221– 258. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1542-4774.2012.01101.x.
- _, James Poterba, and Casey Rothschild, "Redistribution by insurance market regulation: Analyzing a ban on gender-based retirement annuities," *Journal of Financial Economics*, January 2009, 91 (1), 38–58.
- __, Nathaniel Hendren, and Erzo F. P. Luttmer, "The Value of Medicaid: Interpreting Results from the Oregon Health Insurance Experiment," *Journal of Political Economy*, December 2019, *127* (6), 2836–2874. Publisher: The University of Chicago Press.
- Fleurbaey, Marc, "Health, Equity and Social Welfare," *Annales d'conomie et de Statistique*, 2006, (83/84), 21–59. Publisher: [GENES, ADRES].
- Fogel, Robert W. and Chulhee Lee, "Who Gets Health Care?," July 2003.

- Garber, Alan M. and Charles E. Phelps, "Economic foundations of cost-effectiveness analysis," *Journal of Health Economics*, February 1997, *16* (1), 1–31.
- Hendren, Nathaniel, "Measuring economic efficiency using inverse-optimum weights," *Journal* of Public Economics, July 2020, 187, 104198.
- Jacobs, Bas, Egbert L. W. Jongen, and Floris T. Zoutman, "Revealed social preferences of Dutch political parties," *Journal of Public Economics*, December 2017, *156*, 81–100.
- _ , _ , and _ , "Revealed social preferences of Dutch political parties," *Journal of Public Economics*, December 2017, *156*, 81–100.
- **Jesus, Seade**, "On the optimal taxation of multidimensional consumers," Technical Report 7921, CEPREMAP 1979. Publication Title: CEPREMAP Working Papers (Couverture Orange).
- Keeler, Emmett B. and John E. Rolph, "The demand for episodes of treatment in the health insurance experiment," *Journal of Health Economics*, December 1988, 7 (4), 337–367.
- Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez, "The Optimal Income Taxation of Couples," *Econometrica*, 2009, 77 (2), 537–560. Publisher: [Wiley, The Econometric Society].
- Landais, Camille and Johannes Spinnewijn, "The Value of Unemployment Insurance," *The Review of Economic Studies*, November 2021, 88 (6), 3041–3085.
- _ and _ , "The Value of Unemployment Insurance," *The Review of Economic Studies*, November 2021, 88 (6), 3041–3085.
- Lockwood, Benjamin B. and Matthew Weinzierl, "Positive and normative judgments implicit in U.S. tax policy, and the costs of unequal growth and recessions," *Journal of Monetary Economics*, 2016, 77 (C), 30–47. Publisher: Elsevier.
- and __, "Positive and normative judgments implicit in U.S. tax policy, and the costs of unequal growth and recessions," *Journal of Monetary Economics*, February 2016, 77, 30–47.
- Mankiw, N. Gregory, Matthew Weinzierl, and Danny Yagan, "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, December 2009, *23* (4), 147–174.
- Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz, "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment," *The American Economic Review*, 1987, 77 (3), 251–277. Publisher: American Economic Association.
- McClellan, Mark and Jonathan Skinner, "Medicare Reform: Who Pays And Who Benefits?," *Health Affairs*, June 2018. Publisher: Project HOPE The People-to-People Health Foundation, Inc.
- Mirrlees, J. A., "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 1971, *38* (2), 175–208. Publisher: [Oxford University Press, Review of Economic Studies, Ltd.].
- Newhouse, Joseph P., Free for All?, Harvard University Press, February 1996.

- and Charles E. Phelps, "New Estimates of Price and Income Elasticities of Medical Care Services," in "The Role of Health Insurance in the Health Services Sector," NBER, 1976, pp. 261–320.
- Newhouse, Joseph, Robert Brook, Naihua Duan, Emmett Keeler, Willard Manning, M Marquis, Carl Morris, Charles Phelps, and John Rolph, "Attrition in the RAND Health Insurance Experiment: A Response to Nyman," *Journal of health politics, policy and law*, May 2008, *33*, 295–308; discussion 309.
- Nyman, John A., "Health Plan Switching and Attrition Bias in the RAND Health Insurance Experiment," *Journal of Health Politics, Policy and Law*, April 2008, *33* (2), 309–317.
- **Pauly, Mark V., Bradley Herring, and David Song**, "Tax Credits, the Distribution of Subsidized Health Insurance Premiums, and the Uninsured," September 2001.
- **Piketty, Thomas and Emmanuel Saez**, "A Theory of Optimal Inheritance Taxation," *Econometrica*, 2013, *81* (5), 1851–1886. Publisher: [Wiley, Econometric Society].
- _, _, and Stefanie Stantcheva, "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities," *American Economic Journal: Economic Policy*, February 2014, 6 (1), 230–271. *Project MUSE Extending the Theory to Meet the Practice of Insurance*
- *Project MUSE Extending the Theory to Meet the Practice of Insurance.* relationship of of two random variables with proportional covariance - Google Search
- relationship of of two random variables with proportional covariance Google Search.
- Rochet, Jean-Charles and Philippe Chone, "Ironing, sweeping, and multidimensional screening," *Econometrica*, July 1998, *66* (4), 783–826. Num Pages: 44 Place: Evanston, United Kingdom Publisher: Blackwell Publishing Ltd.
- Saez, Emmanuel, "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 2001, 68 (1), 205–229. Publisher: Oxford University Press.
- _, "Direct or indirect tax instruments for redistribution: short-run versus long-run," *Journal of Public Economics*, March 2004, 88 (3-4), 503–518.
- _, "Optimal progressive capital income taxes in the infinite horizon model," *Journal of Public Economics*, January 2013, 97, 61–74.
- and Stefanie Stantcheva, "Generalized Social Marginal Welfare Weights for Optimal Tax Theory," American Economic Review, January 2016, 106 (1), 24–45.
- and __, "A simpler theory of optimal capital taxation," *Journal of Public Economics*, June 2018, 162, 120–142.
- Shields-Zeeman, Laura and Filip Smit, "The impact of income on mental health," *The Lancet Public Health*, June 2022, 7 (6), e486–e487. Publisher: Elsevier.
- **Skinner, Jonathan and Weiping Zhou**, "The Measurement and Evolution of Health Inequality: Evidence from the U.S. Medicare Population," October 2004.
- _ and _ , "The Measurement and Evolution of Health Inequality: Evidence from the U.S. Medicare Population," October 2004.

Stantcheva, Stefanie, "Understanding economic policies: What do people know and learn?," 2020.

- Vera-Hernndez, Marcos, "Structural Estimation of a Principal-Agent Model: Moral Hazard in Medical Insurance," *The RAND Journal of Economics*, 2003, 34 (4), 670–693. Publisher: [RAND Corporation, Wiley].
- Wadhera, Rishi K., Yun Wang, Jose F. Figueroa, Francesca Dominici, Robert W. Yeh, and Karen E. Joynt Maddox, "Mortality and Hospitalizations for Dually Enrolled and Nondually Enrolled Medicare Beneficiaries Aged 65 Years or Older, 2004 to 2017," *JAMA*, March 2020, *323* (10), 961–969.

Appendix

Proof of Proposition 2

Proposition 2 The optimal "low-income eligibility" threshold \hat{z} is characterized by by:

$$\bar{g}_m(\hat{z}) = (\eta_m(s_H, s_L - s_H | \hat{z}) + 1) \quad \text{with} \quad \bar{g}_m(\hat{z}) = \frac{\int_{\mathcal{M}} g_i H_i \cdot m_i \, dF_{m,z}(m_i | \hat{z})}{\lambda \int_{\mathcal{M}} m_i dF_{z,m}(m_i | \hat{z})},$$

and where $\eta_m(s_H, s_L - s_H | \hat{z})$ denotes the pairwise elasticity between s_L and s_H .

Proof. To begin, consider the effect of raising \hat{z} by $\epsilon > 0$ on welfare. Denote the original policy by $P = (R, \tau, \hat{z}, s_L, s_H)$, and the reformed policy by P', where $P' = (R, \tau, \hat{z}', s_L, s_H)$, where the only difference between P and P' comes from changing the \hat{z} to \hat{z}' and all other policy instruments are held fixed. Consider reforming the policy by lowering the eligibility threshold from \hat{z} to $\hat{z} + \epsilon$ for $\epsilon > 0$. The welfare effects from this policy reform are given by

$$W(P'|\hat{z}+\epsilon) - W(P|\hat{z}) = \int_{\hat{z}}^{\hat{z}+\epsilon} \int_{\mathcal{M}} g_i(\underbrace{U_i(s_L)}_{\text{utility at the}} - \underbrace{U_i(s_H)}_{\text{utility at}}) dF_{z,m}(m_i|z_i).$$

If $s_L > s_H$ and utility is concave in s, then by Jensen's inequality we can say that the aggregate utility gain is bounded above and below by

$$(s_L - s_H) \int_{\hat{z}}^{\hat{z} + \epsilon} \int_{\mathcal{M}} g_i \left. \frac{dU_i}{ds} \right|_{s_L} dF_{z,m}(m_i|z_i) \le W(P'|\hat{z} + \epsilon) - W(P|\hat{z}) \\ \le (s_L - s_H) \int_{\hat{z}}^{\hat{z} + \epsilon} \int_{\mathcal{M}} g_i \left. \frac{dU_i}{ds} \right|_{s_H} dF_{z,m}(m_i|z_i) \le W(P'|\hat{z} + \epsilon) - W(P|\hat{z}) \\ \le (s_L - s_H) \int_{\hat{z}}^{\hat{z} + \epsilon} \int_{\mathcal{M}} g_i \left. \frac{dU_i}{ds} \right|_{s_H} dF_{z,m}(m_i|z_i) \le W(P'|\hat{z} + \epsilon) - W(P|\hat{z})$$

The effect from the policy reform on the budget is given by

$$B(P'|\hat{z}+\epsilon) - B(P|\hat{z}) = \underbrace{-s_L \int_{\hat{z}}^{\hat{z}+\epsilon} \int_{\mathcal{M}} m_i(s_L) dF_{z,m}(m_i|z_i) dF_z(z_i) + s_H \int_{\hat{z}}^{\hat{z}+\epsilon} \int_{\mathcal{M}} m_i(s_H) dF_{z,m}(m_i|z_i) dF_z(z_i)}_{\hat{z}-\mathcal{M}}$$

counterfactual change in insurnace expenditures from shifting people at threshold to the Medicaid rate

We can rearrange and divide everything by the average spending of individuals .

$$B(P'|\hat{z}+\epsilon) - B(P|\hat{z}) = -(s_L - s_H) \left(\eta_m(s_H, s_L - s_H|\Delta \hat{z}_+) + 1\right) \int_{\hat{z}}^{\hat{z}+\epsilon} \int_{\mathcal{M}} m_i(s_H) dF_{z,m}(m_i|z_i) dF_z(z_i)$$

for $\epsilon > 0$, where $\eta_m(s_H, s_L - s_H | \Delta \hat{z}) = \frac{\int_{\hat{z}}^{\hat{z}+\epsilon} \int_{\mathcal{M}} m_i(s_L) dF_{z,m}(m_i|z_i) - \int_{\mathcal{M}} m_i(s_H) dF_{z,m}(m_i|z_i) dF_{z}(z_i)}{\int_{\hat{z}}^{\hat{z}+\epsilon} \int_{\mathcal{M}} m_i(s_H) dF_{z,m}(m_i|z_i) dF_{z}(z_i) \cdot (s_L - s_H)/s_L}$ denotes

the average elasticity between coinsurance subsidies s_L and s_H , averaged over individuals that have

incomes $z_i \in \Delta \hat{z}$, where $\Delta \hat{z}_+$ denotes the range of incomes that are newly eligible for Medicaid under the policy reform, $\Delta \hat{z}_+ = [\hat{z}, \hat{z} + \epsilon]$.

Say the marginal cost of public funds is λ . If the upper bound on the welfare gains is equal to the cost of the policy reform (which will be the budget effect times λ), then the planner will weakly prefer the original policy P to the reformed policy P'. Equating the upper bound of the welfare effect to the budget effect, rearranging terms, and substituting in for the elasticity, gives us the following optimality condition

$$\underbrace{(s_L - s_H) \int\limits_{\hat{z}}^{\hat{z} + \epsilon} \int\limits_{\mathcal{M}} g_i \left. \frac{dU_i}{ds} \right|_{s_H} dF_{z,m}(m_i|z_i)}_{z_H = \lambda(s_L - s_H) \left(\eta_m(s_H, s_L - s_H|\Delta \hat{z}) + 1\right) \int\limits_{\hat{z}}^{\hat{z} + \epsilon} \int\limits_{\mathcal{M}} m_i(s_H) dF_{z,m}(m_i|z_i) dF_z(z_i)}_{z_H = \lambda(s_L - s_H) \left(\eta_m(s_H, s_L - s_H|\Delta \hat{z}) + 1\right) \int\limits_{\hat{z}}^{\hat{z} + \epsilon} \int\limits_{\mathcal{M}} m_i(s_H) dF_z(z_i) dF_z(z_i)}_{z_H = \lambda(s_H - s_H) \left(\eta_m(s_H, s_L - s_H|\Delta \hat{z}) + 1\right) \int\limits_{\hat{z}}^{\hat{z} + \epsilon} \int\limits_{\mathcal{M}} m_i(s_H) dF_z(z_i) dF_z(z_$$

upper bound on welfare gains from raising \hat{z}

cost of policy reform

Dividing both sides by $(s_L - s_H)$, rearranging, and taking $\epsilon \to 0$ results in

$$\frac{\int\limits_{\mathcal{M}} g_i \left. \frac{dU_i}{ds} \right|_{s_H} dF_{z,m}(m_i|\hat{z})}{\lambda \int\limits_{\mathcal{M}} m_i(s_H) dF_{z,m}(m_i|\hat{z})} = \left(\eta_m(s_H, s_L - s_H|\Delta\hat{z}) + 1\right).$$

The marginal effect of s on money-metric utility, U_i , is

$$\int_{\mathcal{M}} g_i \frac{dU_i}{ds} dF_{m,z}(m_i|\hat{z}) = \frac{1}{u'} \int_{\mathcal{M}} g_i \underbrace{\left(-u' \cdot (1-s) \frac{dm_{it}}{ds} H_{it} + u \frac{dH_{it}}{dm_{it}} \frac{dm_{it}}{ds} + \frac{dEV_{i,t+1}}{dm_{it}} \frac{dm_{it}}{ds}}_{=0 \text{ by the envelope theorem}} + u' \cdot m_{it} \cdot H_i \, dF_{m,z}(m_i|\hat{z}),$$

$$= \int_{\mathcal{M}} g_i \cdot m_{it} \cdot H_{it}(m_{it}, \theta_i, a_t) dF_{m,z}(m_i|\hat{z}),$$

and substituting in results in the desired expression.

Detailed Lagrangian Derivation of the Optimal Policy

Denote a particular policy by P, where P consists of an income tax schedule and a health insurance schedule. Suppose the government assigns welfare weights to individuals based on observable characteristics of individuals valued by society, such as health state and income. Suppose that the welfare weights are defined over characteristics of individuals i at the optimal policy, P. That is, while the characteristics of individuals may be endogenous to the policy, (e.g. health state may be a function of health insurance), the social welfare weights are fixed, and defined at a particular vector of individual characteristics that correspond to the equilibrium values *at the* optimal policy, P.

The government sets policy by maximizing an aggregate objective of individual's money-metric utility, weighted by a particular set of social welfare weights. Denoting individuals by i, the government chooses a policy P optimally to maximize

$$W(\tilde{P}|P) = \int_{i} g_i(H_i, z_i|P) U_i(\tilde{P}) di,$$

where \tilde{P} is an arbitrary policy, $g_i(P)$ the generalized marginal social welfare weight assigned to individual *i* at the optimal policy *P*, and $U_i(\tilde{P})$ the money-metric utility of individual *i* at any arbitrary policy \tilde{P} .

Consider the case in which the planner chooses an optimal policy $P_t = (s_H, s_L, \hat{z}, \tau, R)$ at a fixed point in time t to maximize aggregate welfare of the cross-section of individuals.

$$P_t \in \underset{\tilde{P}_t}{\arg \max} \int_i g_i(P) \frac{1}{u'} V_{i,t}(\tilde{P}_t) di$$

where

$$V_{i,t} = u(c_{it} - v_i(z_{it}))H_{it}(m_{it}, \theta_{it}, a_t) + EV_{i,t+1}(m_{it})$$

and there are no savings so $c_{it} = R_t + (1 - \tau)z_{it} - (1 - s)m_{it}$, and the continuation value depends only on current medical spending,

$$EV_{i,t+1} = \sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^{t+k} u(c_{i,t+k} - v_i(z_{i,t+k}))S(t+k|a_{t+k}, m_{it})H_{it}(m_{it}, \theta_{it}, a_t).$$

The Lagrangian formulation of the planner's problem is given by:

$$\mathcal{L} = \int_{i} \frac{g_{i}}{u'} (u(\underbrace{R + (1 - \tau)z_{i} - (1 - s(z))m_{i}}_{=c_{i}} - v_{i}(z_{i}))H_{i}(m_{i}, \theta_{it}, a_{t}) + EV_{i,+1}(m_{i})) di$$
$$+ \lambda \left(\tau \int_{i} z_{i} di - R - \int_{0}^{\hat{z}} \int_{\mathcal{M}} s_{L}m_{i} dF_{zm}(m_{i}|z_{i})dF_{z}(z_{i}) - \int_{\hat{z}}^{Z} \int_{\mathcal{M}} s_{H}m_{i} dF_{zm}(m_{i}|z_{i})dF_{z}(z_{i})\right)$$

subject to $s, \tau \in [0, 1]$

The optimality condition for the payroll tax is

$$\begin{aligned} \frac{d\mathcal{L}}{d(1-\tau)} &= \int_{i} \frac{g_{i}}{u'} H_{it}(m_{it}, \theta_{it}, a_{t}) \cdot u'(c_{it} - v_{i}(z_{it})) \left(\underbrace{(1-\tau)\frac{dz_{i}}{d(1-\tau)} - v'_{i}(z_{i})\frac{dz_{i}}{d(1-\tau)}}_{=0} + z_{i}\right) di \\ &+ \lambda \underbrace{\left(\tau \int_{i} \frac{dz_{i}}{d(1-\tau)} di - \int_{i} z_{i} di\right)}_{=\bar{z}(\frac{\tau}{1-\tau}\xi_{z}-1)}, \end{aligned}$$

and setting it equal to zero implies that $\left(\frac{\tau}{1-\tau}\xi_z - 1\right) = \frac{\int_i g_i H_{it}(m_{it},\theta_{it},a_t)\cdot z_i \, di}{\lambda \cdot \int_i z_i di} = \bar{g}_z$ at the optimum. The optimality condition for the two coinsurance rates is given by

$$\begin{split} \frac{d\mathcal{L}}{d(1-s)}\Big|_{z_i\in\mathcal{Z}'} &= \int_{\mathcal{Z}'} \int_{\mathcal{M}} \frac{g_i}{u'} (\underbrace{-u'\cdot(1-s)\frac{dm_{it}}{d(1-s)}H_{it} + u\frac{dH_{it}}{dm_{it}}\frac{dm_{it}}{d(1-s)} + \frac{dEV_{i,t+1}}{dm_{it}}\frac{dm_{it}}{ds}}{-u'\cdot m_{it}H_{it}dF_{m,z}(m_i|z_i)dF_z(z_i)} \\ &= 0 \text{ by the envelope theorem} \\ &+ \lambda \left(\underbrace{\frac{s}{(1-s)}}_{\eta_m(s|\mathcal{Z}')\text{ the elasticity of medical spending among group } \mathcal{Z}'}_{\eta_m(s|\mathcal{Z}')\text{ the elasticity of medical spending among group } \mathcal{Z}'} + 1 \right) \underbrace{\int_{\mathcal{Z}'} \int_{\mathcal{M}} m_i dF_{m,z}(m_i|z_i) dF_z(z_i)}_{\bar{m}(\mathcal{Z}') \text{ average medical spending among } \mathcal{Z}'} \\ \end{split}$$

Setting it equal to zero implies that $\left(\frac{s}{(1-s)}\eta_m(s|\mathcal{Z}')+1\right) = \frac{\int \int g_i \cdot m_{it} \cdot H_{it} dF_{m,z}(m_i|z_i) dF_z(z_i)}{\lambda \cdot \int \int m_{it} \cdot dF_{m,z}(m_i|z_i) dF_z(z_i)}$ at the optimum.

The optimality condition for the transfer, R, pins down the multiplier, λ .

$$\frac{d\mathcal{L}}{dR} = \int_{i} \frac{g_{i}}{u'} H_{it}(m_{it}, \theta_{it}, a_{t}) \cdot u'(c_{it} - v_{i}(z_{it})) \, di - \lambda \qquad \Longrightarrow \ \lambda = \int_{i} g_{i} H_{it}(m_{it}, \theta_{it}, a_{t}) \, di$$

Finally, the optimality condition for the safety net eligibility threshold, \hat{z} , is given by:

$$\begin{aligned} \frac{d\mathcal{L}}{d\hat{z}} &= f(\hat{z}) \int_{\mathcal{M}|\hat{z}} g_i(u(R+(1-\tau)z_i - (1-s_L)m_i - v_i(z_i))H_i(m_i,\theta_{it},a_t) + EV_{i,+1}(m_i))dF_{zm}(m_i|\hat{z}) \\ &- f(\hat{z}) \int_{\mathcal{M}|\hat{z}} g_i(u(R+(1-\tau)z_i - (1-s_H)m_i - v_i(z_i))H_i(m_i,\theta_{it},a_t) + EV_{i,+1}(m_i))dF_{zm}(m_i|\hat{z}) \\ &+ \lambda \left(-\int_{\mathcal{M}} s_L m_i \, dF_{zm}(m_i|\hat{z}) + \int_{\mathcal{M}} s_H m_i \, dF_{zm}(m_i|\hat{z})) \right) f(\hat{z}). \end{aligned}$$