Analyzing the Effects of Insuring Health Risks:*

On the Trade-off between Short-Run Insurance Benefits vs. Long-Run Incentive Costs

Harold L. Cole
University of Pennsylvania and NBER

Soojin Kim
Purdue University

Dirk Krueger
University of Pennsylvania, CEPR, CFS, NBER and Netspar

February 9, 2018

Abstract

This paper quantitatively evaluates the trade-off between the provision of health-related social insurance and the incentives to maintain good health through costly investments. To do so, we construct and estimate a dynamic model of health investments and health insurance in which the cross-sectional health distribution evolves endogenously and is shaped by labor market and health insurance policies. A *no wage discrimination law* in the labor market limits the extent to which wages can depend on the health status of a worker, and a *no prior conditions law* outlaws higher insurance premia for individuals with worse health status. In the model, the static gains from better insurance against poor health induced by these policies are traded off against their adverse dynamic incentive effects on household efforts to lead a healthy life. In our quantitative analysis, we find that it is optimal to insure 80% of labor market-related income risk (70% if a no prior conditions law is also present). Providing full insurance is strongly suboptimal, however, since at high levels of consumption insurance, the negative dynamic incentive effects on health effort and thus the population health distribution in the long run start to dominate the short-run consumption insurance gains.

JEL Codes: E61, H31, I18

Keywords: Health Risks, Social Insurance, Health Effort Choices

1 Introduction

In this paper we study the impact of social insurance policies aimed at reducing a household's exposure to health-related risk in health care and labor markets. We model and quantitatively examine the trade-off between the benefits of greater insurance against health risks in these two contexts, and the resulting reduction in the incentives of households to maintain their health. In many countries, government policies or regulations restrict the extent to which health insurance companies can condition health insurance premia on pre-existing differences in the health conditions of individuals. We will call such legislation that insures individuals against health-dependent expenditure risk on health insurance premia a no prior conditions law. For example, in Switzerland health insurance is compulsory and premia charged by private insurance companies cannot depend on the health status of the insured. In the U.S., the Patient Protection and Affordable Care Act (ACA henceforth) contains a provision that requires health insurers to offer the same insurance premium to all applicants of the same age and location without regard to gender or pre-existing health conditions.

Similarly, in the context of the labor market, across a wide set of countries, regulations are in place that limit the extent to which employers can condition a worker's compensation on (the change in) her health

*We thank Pietro Biroli, Flavio Cunha, Hanming Fang, Gautam Gowrisankaran, Roozbeh Hosseini, Iourii Manovskii, Michele Tertilt, Ken Wolpin, four referees and participants at various institutions for extremely helpful comments, and we gratefully acknowledge support from the NSF (grant SES-1326781).

status. Examples of such regulations, which we label wage non-discrimination laws in this paper, include special employment protection, job quotas and wage subsidies for the disabled in Germany. For the U.S., the Americans with Disabilities Act Amendments Act (ADAAA) of 2008 significantly broadened the restrictions on employers imposed by the Americans with Disabilities Act (ADA) from 1990.[1]

The size of health-related risks and the importance of potential incentive effects from no prior conditions laws and wage non-discrimination legislation rest on three empirical observations that we document for the U.S. in section 2. First, a better health status increases a worker's productivity and thus her labor earnings to a sizable degree. Second, while somewhat less important than the impact on earnings, a good health status reduces the chances of getting acutely sick and thus reduces health expenditure risk to a significant degree. Third, households can affect the evolution of their health status over time by taking costly (in terms of resources or utility) actions such as exercising and abstaining from smoking, and over time, this has an appreciable impact on the health distribution of the population. Based on these observations, the main argument of the paper is that it is crucial to study no prior conditions laws in the health insurance market and no wage discrimination laws in the labor market *jointly*, since their interaction provides social insurance against health-related risks, but negatively impacts health effort incentives.

To quantitatively analyze the impact of these policies, we construct a dynamic life-cycle model with endogenous and stochastically evolving health. A person's health status is an individual state variable that determines both a household's productivity at work and the likelihood that it is subject to an adverse health-related shock. This shock in turn affects productivity and can be offset by medical expenditures for which individuals can purchase health insurance.[2] Health status itself is persistent and changes stochastically over time, and its evolution is affected by the household's effort choice to maintain its health. Since health-related social insurance policies reduce an individual's economic incentives to maintain her health, a trade-off between the provision of social insurance and private incentives emerges, rendering the adoption of these policies a non-trivial policy design question. In addition, the endogenous demand for medical expenditures and thus private insurance contracts respond to the public policy regime.

We use the model as a theoretical and quantitative laboratory for the study of no prior conditions and no wage discrimination social insurance policies. In order to isolate the social insurance benefits and incentive costs of both policies, we first study the static and the dynamic impact of idealized versions of both policies that fully insure, at no administrative costs, the health-related income and expenditure risks, respectively. The static analysis holds the population health distribution fixed and focuses on the equilibrium health insurance contract and the provision of consumption insurance against adverse health status by the policies. In contrast, the key aspect of the dynamic analysis is the impact the policies have on individuals' incentives to maintain their health, and the interaction this creates between the health distribution of the population and the costs of health insurance and the productivity of the workforce. After evaluating the idealized versions of these policies, we also assess the insurance benefits and incentive costs of no wage discrimination laws that are only *partially effective* in insuring health-related income risk, and might require resource costs to be enforced. This last analysis reflects our view that, whereas strong forms of no prior conditions laws that rule out health-condition-based insurance premia are ubiquitous across countries,[3] the degree of labor income insurance through no wage discrimination type legislation varies significantly across countries and

[1] For a broader overview over no wage discrimination and no prior conditions legislation across countries, see Table 11 in the appendix. Although legislations often refer to disabled workers and the precise definition of a disability differs across countries, the OECD (2010) shows that disability is not a marginal phenomenon, afflicting on average across OECD countries 14% of all workers, rising above 20% in some countries. In addition, many countries have strengthened labor market integration policies for disabled workers in the recent decade, often broadening the definition of a disability, implying that more social insurance is provided for unhealthy workers. The most common definition of a disability is a "chronic health problem for at least six months limiting daily activities."For example, under the ADAAA in the U.S. individuals whose cancer is in remission or whose diabetes is controlled by medication are still considered disabled under the act. Recently, conditions like morbid obesity and alcoholism have been deemed disabilities. See https://www.ada.gov/regs2016/final rule adaaa.html for the regulatory assessment by the Justice Department and http://www.nytimes.com/roomfordebate/2015/07/26/the-americans-with-disabilities-act-25-years-later/a-bright-spot-in-the-law-including-obesity and https://www.ada.gov/employmt.htm for a discussion of the broadening of the notion of disability under the ADAAA, which replaced the largely ineffective original ADA (see, e.g., Befort, 2013)

[2] We also model catastrophic health shocks that require non-discretionary health expenditures to avoid death.

[3] In the U.S. under the ACA, since 2010 children below the age of 19 cannot be excluded from their parents' health insurance policy or denied treatment for pre-existing conditions. Since 2014, this last restriction applies to adults as well. Moreover, insurance companies are no longer able to use health status to determine eligibility, benefits or premia, and are prevented from limiting lifetime or annual benefits.

over time (as in the U.S. when the ADAAA replaced the largely ineffective original ADA) and is unlikely as complete and cost-effective as our idealized version of the no wage discrimination law models it.[4]

To empirically implement our quantitative analysis, we first estimate and calibrate the model to U.S. PSID and MEPS data to match key statistics on labor earnings, medical expenditures and physical exercise levels. We then use the model as a quantitative laboratory to evaluate the consequences of the different policy options. Although, in the U.S. context, we think of the ACA as providing the no prior conditions legislation and the ADAAA as the broad motivation for the no wage discrimination policy, our focus is on the specific role of these classes of policies in providing social insurance against health-related income and health insurance premium risk as well as their negative incentive effects. Ours is clearly *not* intended to be a comprehensive study of all other aspects and provisions of the ACA or health-related labor market policies such as the ADAAA. At the same time, given the prevalence of similar policies in other countries, the relevance of our analysis is not limited to the U.S. context either.

Our quantitative analysis reveals that it is optimal, from an ex-ante lifetime utility perspective, to insure 80% of labor market-related income risk (and 70%, if a no prior conditions law is also present). On the other hand, providing full insurance, by combining a fully effective no wage discrimination and a no prior conditions law, is strongly suboptimal, since at high levels of consumption insurance, the negative dynamic incentive effects on health effort and thus the population health distribution in the long run start to dominate the short-run, static consumption insurance gains. Thus, overall, health-related risks in the labor and health insurance market call for strong social insurance, and both policies studied in this paper are welfare improving relative to the competitive equilibrium. The optimal extent of insurance remains large even when we consider empirically plausible resource costs associated with the implementation of the no-wage-discrimination policy. However, even without any resource costs, complete insurance is never optimal.

At the broadest level, our paper positively and normatively evaluates the trade-off between the provision of income *insurance* and the distortion of *incentives* induced by social insurance policies, but shifts the focus to health-related (insurance) policies. On this general level, our work therefore builds upon the large literature assessing this trade-off for other social insurance policies, such as (among others) unemployment insurance (e.g., Shavell and Weiss (1979), and Hopenhayn and Nicolini (1997)), disability insurance (e.g., Golosov and Tsyvinski (2007), and Low and Pistaferri (2015)) and progressive income taxation (e.g., Mirrlees (1971)) and the large literatures that followed these contributions.

The health-related social insurance policies we study provide income and thus consumption insurance benefits in much the same way as in the general social insurance literature cited above. On the other hand, the dynamic health effort incentive effects are specific to the health policies we model. They impact the future evolution of individual health status, and through it, future earnings capabilities, and therefore the distribution of income and health in the economy as a whole. The paper thus connects and contributes to three specific literatures at the intersection of quantitative macroeconomics and health economics. First, on the modelling side, our structure with endogenous effort choice and the dynamic and endogenous evolution of health status builds upon the strand of the quantitative macroeconomics literature that has endogenized the evolution of health status over the life cycle. Second, it builds upon the empirical literature investigating the impact of health or diseases on earnings, and the determinants of the dynamics of health, since our theoretical mechanism crucially relies on these links. Finally, since in the U.S., the economy to which we calibrate our model, the social insurance we model is motivated by selected provisions in the ACA and the ADAAA, we contribute to the literature evaluating these health-related policies.

Starting with the seminal works of Grossman (1972) and Ehrlich and Becker (1972), the *first* literature starts with the insight that health is in part an investment good whose evolution can be actively impacted by investing effort or resources. Our model, in which dynamic (and stochastic) health transitions can be influenced by costly health effort, directly builds upon this tradition. A recent model-based quantitative

[4]The ADAAA requires that if a worker with a disability can meet the same production standards for essential job functions as a worker without a disability with the help of a reasonable accommodation, then this worker must receive the same wage as the worker without the disability. Importantly, the employer has to pay for the accommodation, unless it would cause undue hardship for the employer. In the absence of the ADAAA, the cost of the accommodation may be reflected in the wage of the worker. To determine empirically the exact extent to which the ADAAA provides income insurance is difficult. Therefore mapping the model policies into the U.S. context, we initially focus on the extreme case in which ADAAA eliminates all health-related income variation, and then turn to policies with less than perfect insurance. We thank a conscientious referee for detailed and extremely helpful discussions about the scope of the ADAAA and its effects on wages.

literature has used related dynamic models with dynamic health updates to study the macroeconomic and distributional implications of health, health insurance and health care policy reforms.[5] Within this first literature, the most closely related papers are Brügemann and Manovskii (2010) and Jung and Tran (2016), since they also study the effects of the ACA on health insurance coverage and macroeconomic aggregates. Aizawa and Fang (2015) extend the labor market mobility model with the employer-sponsored health insurance constructed by Dey and Flinn (2005) to study the effects of the ACA. As with Brügemann and Manovskii (2010) and Jung and Tran (2016), neither of these last papers is concerned with the incentive effects on *health efforts* and thus health transitions induced by regulation in *both* the labor and the health insurance markets (and crucially, *their interaction*) that we formalize in our model.

The *second* empirical strand of the literature that this paper builds upon studies the impact of general health or specific diseases on earnings, and on the determinants of the dynamics of health. The empirical literature estimating the impact of health on income (see, e.g., Bartel and Taubman (1979), Mitchell and Butler (1986), Cawley (2004) and Currie and Madrian (1999) for a summary) finds a positive impact of health on earnings. Concerning the dynamics of health transitions, Pijoan-Mas and Rios-Rull (2014) document an important impact of socioeconomic status (most importantly, an education dependence that we also permit in our model). Moreover, many studies in economics and the medical literature (e.g., Colman and Dave (2012), Booth et al. (2012)) find that individual behavior is a significant determinant of health status changes over time, which is the key premise of our model.[6] Finally, there is evidence that such health behavior responds to economic incentives. Bhattacharya et al. (2011) employ data from a Rand health insurance experiment to show that access to health insurance leads to increases in body mass and obesity because insurance insulates people from the impact of their excess weight on their medical expenditure costs. Charness and Gneezy (2009) use experimental data to show that individuals' gym attendance responds significantly and persistently to financial incentives.[7]

Finally, in addition to the papers that evaluate the impact of the ACA in the U.S. in structural dynamic models, our study complements a *third* strand of the literature that estimates the effect of the original ADA legislation in the U.S. from 1990 on employment, wages and labor hours of the disabled (see, e.g., Acemoglu and Angrist (2001), and DeLeire (2000, 2001)). Most find that the employment rate of the disabled was lower after the ADA. DeLeire (2001) documents that, compared to 1984, the earnings gap between the disabled and he non-disabled fell significantly in 1993, some of which may be attributable to the ADA. These studies focus on evaluating the impact of the ADA on the labor market performance of the disabled. On the other hand, we interpret the ensuing Amendments Act as a shift in government policies toward providing more insurance against broader health-related risks and we study its impact on he health behavior of workers.

The paper is organized as follows. In section 2 we set out the empirical facts justifying our modelling approach. We describe the model and implementation of the two policies in section 3. The theoretical analysis of the static and dynamic version of the baseline model is contained in section 4. In section 5 we describe how we augment the model to map it into the data, as well as our estimation and calibration procedure. Section 6 presents the main quantitative results of the policy analysis, of both the idealized and the partial insurance versions of the policies. Robustness analyses are contained in section 7, and section 8 concludes. Proofs and details of the quantitative analysis are relegated to the appendix.

2 Motivating Empirical Facts

Our theoretical model is built on three premises: first, that a good health status increases a worker's productivity and thus labor earnings; second, that a good health status reduces the chances of getting acutely sick and thus reduces expected health expenditures; and third, that workers can affect the dynamic

[5]This literature includes French and Jones (2011), Hall and Jones (2007), De Nardi et al. (2016), Jeske and Kitao (2009), Attanasio et al. (2011), Ales et al. (2014), Braun et al. (2017), Halliday et al. (2016), Hansen et al. (2014), Ozkan (2014), and Pashchenko and Porapakkarm (2013).

[6]The empirical findings on the impact of medical *expenditures* on health outcomes is mixed. Baicker et al. (2013) study the Medicaid expansion in Oregon and find that despite an increased utilization of medical services, there were no significant improvements in health outcomes after 2 years. Recent studies of the ACA (e.g., Barbaresco, et al. (2015), Simon, et al. (2017)) find at most small increases in self-reported health statuses among young (23-25 year olds) and low-income childless adults.

[7]Kowalski (2015) investigates empirically, employing a static model, the trade-off between insurance and the moral hazard effects of health insurance provision on medical spending, finding that the latter outweigh the former.

evolution of their health statuses by taking costly actions. The public policies under study then provide additional social insurance against labor income risk and health expenditure risk, but also have an adverse impact on the dynamic incentives to lead healthy lives. The purpose of the paper is to qualitatively and quantitatively evaluate this trade-off. Before we turn to this task, we first want to document the empirical plausibility of the three basic premises on which our argument is being built. In this section we focus on raw correlations motivating our analysis; when we estimate the model in section 5, we take into account that part of these correlations could be driven by observable and unobservable household heterogeneity.

In Table 1 we use PSID data from 1999 to 2009, sort individuals into four health groups ranging from (self-reported) fair to excellent health and document that *better health is associated with significantly higher labor income*.[8] Labor income is strongly increasing in health: for example, mean (median) labor income among those reporting excellent health is 116% (84%) larger than for individuals with fair health. As documented in section 5, this health-income gap shrinks to 67% when we control for other observable differences across individuals and therefore remains economically very significant.[9]

Table 1: Labor Income by Health Status

Health Status	Labor Income, if positive		
	Mean	St. Dev.	Median
Fair	32,752	29,211	26,483
Good	45,970	46,615	36,665
Very Good	55,541	79,465	41,604
Excellent	70,826	129,021	48,695
All	55,075	867,289	40,797

Table 2: Medical Expenditure by Health Status

	Medical Expenditure		
	Mean	St. Dev.	Median
Fair	5,821	13,043	1,977
Good	2,344	6,118	733
Very Good	1,601	3,861	558
Excellent	1,227	2,872	363
All	2,157	6,172	599

Second, in Table 2 we exploit data from the 1997 to 2002 waves of the Medical Expenditure Panel Survey (MEPS) to document the *negative correlation between health status and health expenditures*. Individuals are asked to self-report their health status on the same scale as in the PSID, and in the table, we display the mean and median health expenditures across health status groups. We use the MEPS for medical expenditures, as this data set reports individual-level medical expenditures, whereas the PSID reports total medical expenditures only at the household level. Also, expenditures in the MEPS include out-of-pocket payments and payments by private insurance, Medicaid, Medicare, and all other sources.

The table shows very significant differences in mean and median health expenditures across individuals with different health statuses: those with fair health spend on average 4.7 times as much as those in the highest health category. In section 5 we document that differences in observable characteristics across health groups are partially responsible for the expenditure gaps, but most of the gap persists after controlling for them (e.g., the mean expenditure ratio between the best and worst health groups drops, from 4.7 to 2.9).

The final, and perhaps most novel, premise of our model is that health status is endogenous and its stochastic evolution can be affected by an individual's effort to lead a healthy life. This premise receives support in the raw data, as Table 3 displays. In this table we summarize, again using data from the PSID, how the dynamics of the health status of individuals is impacted by their effort to lead healthy lives.[10]

The three rows of the table display the share of individuals whose health status, over a six-year interval, either declines, stays the same or improves, such that each row sums to 1. The first two columns document that *individuals with effort above the cross-sectional average are more likely to retain or improve their health statuses*. The differences in health dynamics across two effort groups are significant at a 5% confidence level (using a χ^2 test). The remaining four columns display that the positive association of effort and health transition persists once we control for initial health status.[11]

[8] In the PSID individuals report one of five health statuses. We combine individuals with fair and poor health status to ensure an appropriate sample size in all cells. We use data starting from 1999 because at that date information on individual health efforts become available. All values are in 2000 dollars (throughout the paper).

[9] In our quantitative analysis, and motivated by a subset of the empirical literature on the health-income nexus, we also conduct a robustness analysis with respect to the health-income gradient in which this ratio shrinks to 29%.

[10] Effort is measured as the average frequency of light and heavy physical exercise and the abstention from smoking (number of cigarettes) reported in PSID survey years 1999 through 2011.

[11] Bad health is defined as the lower health statuses (Good, Fair, and Poor); the good health group is composed of the top

Table 3: Effort and Health Dynamics over 6 years

Health	All		Bad Initial Health		Good Initial Health	
	Eff < Avg.	Eff $\geq$ Avg.	Eff < Avg.	Eff $\geq$ Avg.	Eff < Avg.	Eff $\geq$ Avg.
Worsened	0.35	0.30	0.29	0.28	0.39	0.30
Unchanged	0.50	0.52	0.54	0.50	0.48	0.53
Improved	0.15	0.18	0.17	0.22	0.13	0.17

The data presented here suggest a strong role for health status in the determination of income and medical expenditures, and a key role for individual effort in the dynamic updating of health status. We now build a model based on these three premises to evaluate the trade-off between incentive costs and the insurance benefits of social insurance policies in the labor and health insurance market.

3 The Model

Time $t = 0, 1, 2, \ldots T$ is discrete and finite and the economy is populated by a cohort of a continuum of individuals of mass 1. Since we are modeling a given cohort of individuals, we will use time and the age of households interchangeably. We think of T as the end of the working life of the age cohort under study.

3.1 Endowments and Preferences

Households are endowed with one unit of time, which they supply inelastically to the market. They are also endowed with an initial level of health h and we denote by $H = \{h_1, \ldots, h_N\}$ the finite set of possible health levels. Households value current consumption c and dislike the effort e that helps maintain their health. We will assume that their preferences are additively separable over time, and that they discount the future at time discount factor β. We will also assume that preferences are separable between consumption and effort, and that households value consumption according to the common period utility function $u(c)$ and value effort according to the period disutility function $q(e)$.

We will denote the probability distribution over the health status h at the beginning of period t by $\Phi_t(h)$, and denote by $\Phi_0(h)$ the initial distribution over this characteristic.

Assumption 1 *The function u is twice differentiable, strictly increasing and strictly concave. The function q is twice differentiable, strictly increasing, strictly convex, with $q(0) = q'(0) = 0$ and $\lim_{e \to \infty} q'(e) = \infty$.*

3.2 Health and Production Technology

Let ε denote the current health shock. In every period households with current health h with probability $1 - g(h)$ draw a health shock $\varepsilon \in (0, \bar{\varepsilon}]$ that is distributed according to the probability density function $f(\varepsilon)$. With probability $g(h)$ the household draws no shock (that is, $\varepsilon = 0$).

Assumption 2 *f is continuous, g is twice differentiable, strictly increasing and strictly concave.*

An individual's health status evolves stochastically over time, according to the Markov transition function $Q(h', h; e)$, where $e \geq 0$ is the exercise level by the individual. We impose the following assumption on Q.

Assumption 3 *$\frac{\partial Q(h';h,e)}{\partial e}$ is increasing in h' and $\frac{\partial^2 Q(h';h,e)}{\partial e^2}$ is decreasing in h'.*

The first assumption implies first-order stochastic dominance. The second assumption implies that the impact of effort on the transition matrix shrinks as e gets large. Also note that by assumption, health transitions depend on the effort e by individuals, but not on medical expenditures.[12] An individual with health status h, current health shock ε and health expenditures x produces $F(h, \varepsilon - x)$ units of output.

two levels (Very Good and Excellent). Even after controlling for initial health status, the difference between the high and the low effort group is significant for both bad ($\chi^2 = 8.8936, Pr = 0.012$) and good ($\chi^2 = 26.0909, Pr = 0.000$) initial health groups.

[12] Baicker et al. (2013) find that Medicaid expansion led to an increase in health care utilization and detection of diabetes, but without significant improvements in health in the first two years, the horizon of their study.

Assumption 4 *F is continuously differentiable in (h, y), strictly increasing in h, satisfies $F(h, y) = F(h, 0)$ for all $y \leq 0$, and $F_2(h, y) < 0$ and $F_2(h, \bar{\varepsilon}) < -1$. Moreover, $F_{22}(h, y) < 0$ for all $y > 0$ and $F_{12}(h, y) \geq 0$.*

(a) $F(h, \varepsilon - x)$ by Health Status (h)
$h_1 < h_2$

(b) $F(h, \varepsilon - x)$ for Fixed Health Shock (ε)
$h_1 < h_2$ and $\varepsilon_1 < \varepsilon_2$

Figure 1: Production Function by Health Status and Health Shock

The left panel of Figure 1 displays the production function $F(h, .)$, for two different levels of the current health status. Holding health status h constant, output is decreasing in the uncured portion of the health shock $\varepsilon - x$, and the decline is more rapid for lower levels of health ($h_1 < h_2$). The right panel of Figure 1 displays the production function as a function of health expenditures x, for a fixed level of the shock ε, and shows that expenditures x exceeding the health shock ε leave output $F(h, \varepsilon - x)$ unaffected (and thus are suboptimal). Furthermore, a reduction of the shock ε to a lower level, ε_1, shifts the point at which health expenditures x become ineffective to the left.

The assumptions on the production function F imply that health expenditures can offset the impact of a health shock on productivity, but not raise an individual's productivity above what it would be if there had been no shock. In addition, the last assumption on F that $F_{12} \geq 0$ implies that the negative impact of a given net health shock y is lower the healthier a person is.[13] The assumption $F_2(h, \bar{\varepsilon}) < -1$ insures that if an individual is hit by the worst health shock, the cost of treating this health shock, at the margin, is smaller than the positive impact on productivity (output) this treatment has.

Figure 2 summarizes the time line within period t. Households enter the period with individual health status h. In the population of age t, the cross-sectional distribution of health is given by $\Phi_t(h)$. Observing h,

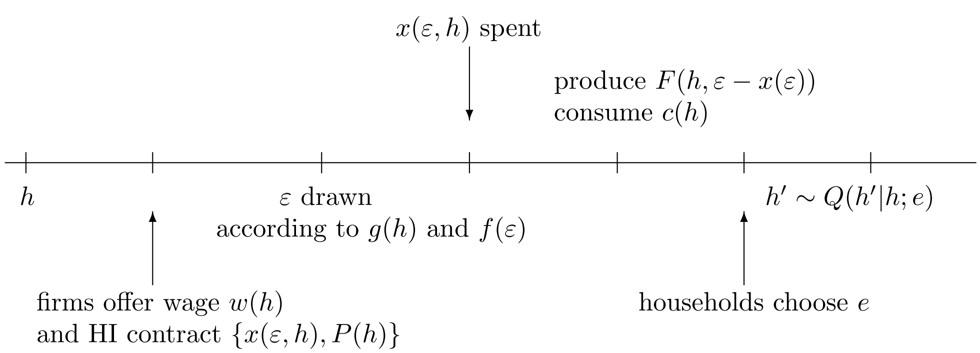


Figure 2: Timing of the Model

[13] This assumption implies that individuals with worse health h will have higher health expenditures, which, as shown above, is consistent with the data. It is also the modelling approach taken by Hugonnier et al. (2013) and Ehrlich and Chuma (1990).

firms then offer wages $w(h)$ and health insurance contracts $\{x(\varepsilon,h),P(h)\}$ to households with health status h which these households accept. Next, the health shock ε is drawn according to the distributions g,f, and then resources according to $x=x(\varepsilon,h)$ are spent on health. Now production and consumption takes place. Finally, at the end of the period, individuals make health effort choices e, and then the new health status h' of a household is drawn according to the health transition function $Q(h'|h;e)$, which, together with $\Phi_t(h)$, determines the new cross-sectional distribution $\Phi_{t+1}(h')$ at the beginning of the next period.

3.3 Market Structure without Government

A large number of production firms in each period compete for workers. Firms observe the health status of a worker h and then, prior to the realization of the health shocks, they compete for workers of type h by offering a wage $w(h)$ that pools the risk of the health shocks and bundle the wage with an associated health insurance contract (specifying health expenditures $x(\varepsilon,h)$ and an insurance premium $P(h)$) that breaks even. Perfect competition for workers of type h requires that the combined wage and health insurance contract maximizes the period utility of the household, subject to the firm breaking even.[14] In the absence of government intervention, a firm specializing in workers of type h then offers a wage $w^{CE}(h)$ (where CE stands for competitive equilibrium) and a health insurance contract $\{x^{CE}(\varepsilon,h),P^{CE}(h)\}$ solving

$$U^{CE}(h) = \max_{w(h),x(\varepsilon,h),P(h)} u\left(w(h)-P(h)\right) \tag{1}$$

$$s.t. \quad P(h) = g(h)x(0,h) + (1-g(h))\int_0^{\bar{\varepsilon}} f(\varepsilon)x(\varepsilon,h)d\varepsilon \tag{2}$$

$$w(h) = g(h)F(h,-x(0,h)) + (1-g(h))\int_0^{\bar{\varepsilon}} f(\varepsilon)F(h,\varepsilon-x(\varepsilon,h))d\varepsilon \tag{3}$$

Note that the source of risk in the competitive equilibrium is the health status risk associated with h. This risk stems from the dependence of both wages $w(h)$ and health insurance premia $P(h)$ on h, and these are exactly the sources of consumption risk that government policies to prevent wage discrimination and prohibit prior health conditions from affecting insurance premia are designed to tackle. Also note that, since $w(h),x(\varepsilon,h),P(h)$ are chosen to maximize $U^{CE}(h)$ for each health status h separately, problem (1)-(3) is equivalent to maximizing $\sum_h \Phi(h)U^{CE}(h)$ subject to constraints (2) and (3).

3.4 Government Policies

We now describe how we model the no prior conditions law and the no wage discrimination legislation.

3.4.1 No Prior Conditions Law

The purpose of a no prior conditions law is to prevent insurance companies from differentially pricing insurance based upon health status.[15] To be successful, regulation must lead to a pooling equilibrium in which all individuals obtain insurance, and obtain it at a price that is independent of h. The best such regulation in addition ensures that the equilibrium health insurance schedule $x(\varepsilon,h)$, given the constraints, is efficient. We now describe the regulations sufficient to achieve this goal.

For the no prior conditions law to be effective, the government must prevent a separating equilibrium in which insurance companies use the health expenditure schedule $x(\varepsilon,h)$ to select their desired health types, given that they are barred from conditioning their premia P on h directly. Therefore, to achieve pooling in the health insurance market, the government must regulate the health expenditure schedule $x(\varepsilon,h)$. To give the legislation the best chance of being beneficial, we assume that the government *regulates the health*

[14]Instead of assuming that firms specialize by hiring only a specific health type of workers h, we could also consider a market structure in which all firms are representative and hire according to the population distribution, and pay with different health h differential wages according to the schedule $w^{CE}(h)$. That is, health variation in wages and variation in hired health types h are perfect substitutes at the level of the individual firm in terms of supporting the competitive equilibrium allocation.

[15]Consistent with this restricted purpose, we will also assume that the government cannot use health insurance policies to offset productivity differences due to other factors such as education. This will prove important in the quantitative section.

expenditure schedule $x(\varepsilon, h)$ *efficiently.* Furthermore, since risk pooling is limited if some household types h do not buy insurance, in the benchmark model we assume that all individuals are *forced* to buy insurance.

Given this regulation and a cross-sectional distribution of workers by health type, Φ, the health insurance premium P charged by competitive firms, given the set of regulations spelled out above, is determined by

$$P = \sum_h \left[g(h)x(0,h) + (1-g(h)) \int f(\varepsilon)x(\varepsilon,h)d\varepsilon \right] \Phi(h) \tag{4}$$

where $x(\varepsilon, h)$ is the expenditure schedule regulated by the government. This schedule is chosen to maximize $\sum_h u(w(h) - P)\Phi(h)$, with wages $w(h)$ determined by (3).

3.4.2 No Wage Discrimination Law

The objective of the government is to prevent workers with a lower health status h, and hence lower productivity, being paid less, to insure them against health status risk. However, if a production firm is penalized for paying workers with low health status h low wages, but not for preferentially hiring workers with a favorable health status h, then a firm can circumvent the no wage discrimination law. Therefore, to be effective, such a law must penalize *both* wage discrimination and hiring discrimination by health status. In the benchmark model we analyze the case where the policy is fully effective (by the threat of punishment) in achieving the goal of preventing differential hiring and compensation.[16]

Under this legislation firms, in their hiring decisions, take as given the economy-wide wage w for a representative worker, which, by perfect competition and zero profits, is given by

$$w = \sum_h \left\{ g(h)F(h, -x(0,h)) + (1-g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) \left[F(h, \varepsilon - x(\varepsilon, h)) \right] d\varepsilon \right\} \Phi(h). \tag{5}$$

Note that this wage depends upon the health expenditure schedule $x(.)$. If firms can discriminate between workers with different health insurance contracts they can effectively circumvent the no wage discrimination law. Therefore, as with the no prior conditions law we need to assume that the government regulates the health insurance market to insure that the no wage discrimination law is fully effective. The regulation again determines the extent of coverage by health type, $x(\varepsilon, h)$, subject to the requirement that the offered health insurance contracts break even, either health type by health type (in the *absence* of a no prior conditions law) or in expectation across health types (in the *presence* of the no prior conditions law). Under the no wage discrimination law, consumption is given by $c(h) = w - P(h)$ in the absence of a no prior conditions law, and by $c(h) = w - P$ in its presence, with $P(h)$ and P in turn given by equations (2) or (4), respectively.[17]

4 Theoretical Analysis of the Model

Given the timing assumptions and the restriction to static health insurance contracts, the model analysis can be separated into a static and a dynamic part. In the static problem, wages and optimal health insurance are chosen for a given distribution Φ. The dynamic problem determines health effort choice e, which leads, through the transition function $Q(h', h; e)$, to a stochastic update of individual health status and thus a new health distribution Φ'. We first analyze the static problem before turning to the dynamics in section 4.2.

4.1 Static Analysis

We now characterize static competitive equilibrium allocations. We first establish the efficient benchmark by analyzing the solution to the social planner problem, in order to highlight the source of inefficiency in the competitive equilibrium (insufficient consumption insurance). The key result is that, statically, the combination of both policies provides full consumption insurance in the regulated market equilibrium, and thus restores the full efficiency of the market outcome.

[16] In Appendix C.2, we discuss the case in which penalties are realized in equilibrium.

[17] Given a health distribution Φ, the efficiently regulated insurance contract $x(\varepsilon, h)$ under a no wage discrimination law maximizes $\sum_h u(w - P(h))\Phi(h)$ subject to (5) and (2) if the no prior conditions law is absent, and subject to (4) instead of (2) if it is present.

4.1.1 Social Planner's Problem

Given an initial cross-sectional distribution over health status in the population $\Phi(h)$, the social planner maximizes ex-ante (prior to the realization of h) household utility (or, as an alternative interpretation, utilitarian social welfare). The planner's problem is given by:

$$U^{SP}(\Phi) = \max_{x(\varepsilon,h),c(\varepsilon,h)\geq 0} \sum_h \left\{ g(h)u(c(0,h)) + (1-g(h)) \int f(\varepsilon)u(c(\varepsilon,h))d\varepsilon \right\} \Phi(h)$$

subject to the economy-wide resource constraint:

$$\begin{aligned} & \sum_h \left\{ g(h)c(0,h) + (1-g(h)) \int f(\varepsilon)c(\varepsilon,h)d\varepsilon + g(h)x(0,h) + (1-g(h)) \int f(\varepsilon)x(\varepsilon,h)d\varepsilon \right\} \Phi(h) \\ \leq & \sum_h \left\{ g(h)F(h,-x(0,h)) + (1-g(h)) \int f(\varepsilon)F(h,\varepsilon - x(\varepsilon,h))d\varepsilon \right\} \Phi(h). \end{aligned}$$

We summarize the solution to the static social planner's problem in the following proposition, whose proof follows directly from the first-order conditions and assumption 4.

Proposition 5 *The solution to the static social planner's problem $\{c^{SP}(\varepsilon,h), x^{SP}(\varepsilon,h)\}_{h\in H}$ satisfies $x^{SP}(\varepsilon,h) = \max\left[0, \varepsilon - \bar{\varepsilon}^{SP}(h)\right]$, where the cutoffs $\{\bar{\varepsilon}^{SP}(h)\}$ satisfy $-F_2(h,\bar{\varepsilon}^{SP}(h)) = 1$, and consumption is given by*

$$c^{SP}(\varepsilon,h) = c^{SP} = \sum_h \left[g(h)F(h,0) + (1-g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) \left[F(h,\varepsilon - x^{SP}(\varepsilon,h)) - x^{SP}(\varepsilon,h) \right] d\varepsilon \right] \Phi(h). \quad (6)$$

The optimal cutoff $\{\bar{\varepsilon}^{SP}(h)\}$ is increasing in h, and strictly so if $F_{12}(h,y) > 0$.

The social planner finds it optimal to provide full consumption insurance not only against adverse health shocks ε, but also against bad health status, as consumption c^{SP} is independent of h. The optimal health expenditure allocation is chosen to maximize the net output contribution $F(h,\varepsilon - x(\varepsilon,h)) - x(\varepsilon,h)$ of a worker with characteristics (ε,h), which gives rise to the cutoff rule and comparative statics in the proposition.

The efficient allocation is depicted in Figure 3. As shown in the proposition, optimal medical expenditures take a cutoff rule: small health shocks $\varepsilon < \bar{\varepsilon}^{SP}(h)$ are not treated at all, but larger shocks are fully treated up to the threshold $\bar{\varepsilon}^{SP}(h)$. The medical expenditures are displayed in Figure 3(b) for two different initial levels of health $h_1 < h_2$: below the h-specific threshold $\bar{\varepsilon}^{SP}(h)$, expenditures are zero, and then rise one for one with the health shock ε. The determination of the threshold itself is displayed in Figure 3(a). It shows that under the assumption that the impact of health shocks on productivity is less severe for healthy households,[18] the equilibrium features better insurance for less healthy households, reflected in a lower threshold for h_1 than for h_2, that is, $\bar{\varepsilon}^{SP}(h_1) < \bar{\varepsilon}^{SP}(h_2)$. The equilibrium health expenditure policy function leads to a net-of-health-treatment production function $F(h,\varepsilon - x^{SP}(\varepsilon,h))$ shown in Figure 3(c).

4.1.2 Competitive Equilibrium without and with Policy

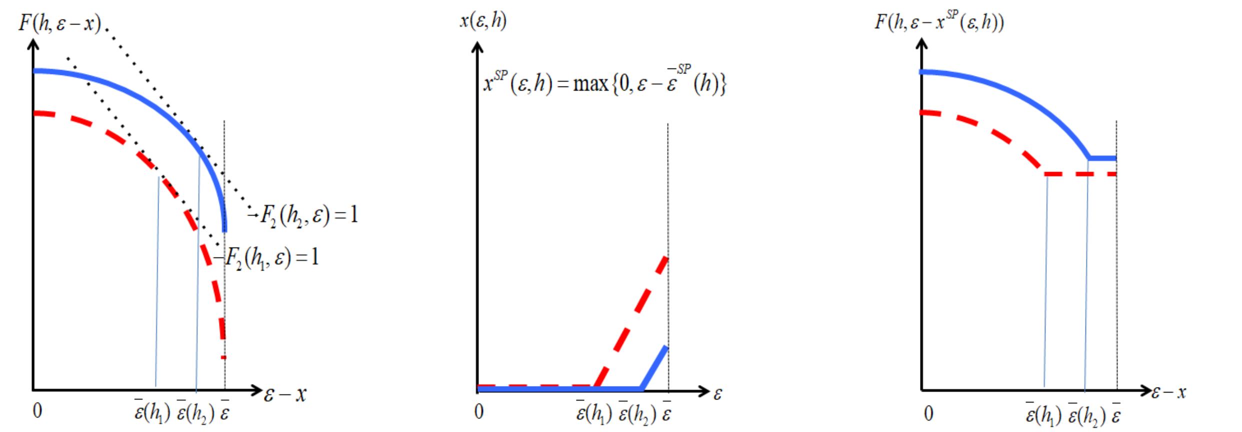
As described in sections 3.3 and 3.4 the equilibrium wage and health insurance contracts solve, depending on the policy regime $i \in \{CE, NP, NW, B\}$ in place

$$U^i(\Phi) = \max_{w(h),x(\varepsilon,h),P(h)} \sum_h u\left(w(h) - P(h)\right)\Phi(h) \quad (7)$$

subject to equations (2) or (4) for premia and equations (3) or (5) for wage.

In the unregulated competitive equilibrium, policy regime $i = CE$, both the health insurance premium and the wage depend on the individual health status h of the worker, as in equations (2) and (3). The no prior conditions legislation ($i = NP$), replaces constraint (2) with (4), and the no wage discrimination law

[18]That is, under the assumption ($F_{12}(h,y) > 0$, reflected as a "more concave" curve for h_1 than for h_2 in Figure 3(a)).



(a) Production Function, $h_1 < h_2$ (b) Optimal Medical Expenditures (c) Net-of-Health-Treatment Production

Figure 3: Optimal Medical Expenditures and Production

$(i = NW)$, (3) with (5). Finally, with both laws in place $(i = B)$, both wages and health insurance premia (and thus individual consumption) are independent of health status h, as constraints (4) and (5) indicate.

We now turn to the theoretical characterization of the competitive equilibrium under the different policy configurations, focusing specifically on the sources of the inefficiency of the laissez-faire competitive equilibrium and how the policies correct them.

Characterization of the Unregulated Equilibrium $(i = CE)$ We now characterize the competitive equilibrium in the absence of policy interventions to isolate the sources of inefficiency in the market solution.

Proposition 6 *The unique equilibrium health insurance contract is given by* $x^{CE}(\varepsilon, h) = \max[0, \varepsilon - \bar{\varepsilon}^{CE}(h)]$, *where the cutoffs satisfy*

$$-F_2(h, \bar{\varepsilon}^{CE}(h)) = 1 \tag{8}$$

and premia are given by equation (2)*. Equilibrium wages are determined by equation* (3) *evaluated at the expenditure profile* $x^{CE}(\varepsilon, h)$*, and consumption is* $c^{CE}(\varepsilon, h) = c^{CE}(h) = w^{CE}(h) - P^{CE}(h)$.

The intuition for the equilibrium health expenditure schedule is simple. For each health status h, the household cares only about net compensation $w(h) - P(h)$. By comparing equations (2) and (3), we observe that for each ε realization, the marginal cost (in terms of the consumption good) of spending an extra unit of x is 1, and the marginal benefit is $-F_2(h, \varepsilon - x(\varepsilon, h))$. The optimal health expenditure schedule equates the two as long as the resulting $x(\varepsilon, h)$ is interior and features $x(\varepsilon, h) = 0$ if for a given ε the benefit of spending the first unit falls short of the cost of 1. Note that the equilibrium health insurance contract has the flavor of deductibles observed in reality (but here the worker pays for $\varepsilon < \bar{\varepsilon}^{CE}(h)$ not with out-of-pocket expenditures but with reduced productivity). It follows directly from Propositions 5 and 6 that the equilibrium health expenditure allocation is efficient: $x^{CE}(\varepsilon, h) = x^{SP}(\varepsilon, h)$ for all (ε, h).

This last observation shows that in the static case the *only* source of inefficiency of the competitive equilibrium stems from the inefficient consumption insurance against adverse prior health conditions h (which was complete in the social planner's solution) since aggregate production and thus consumption are identical in the equilibrium and efficient allocation. The equilibrium allocation of health expenditures is efficient because the firm bundles the determination of wages and the provision of health insurance, and thus internalizes the positive effects of health spending $x(\varepsilon, h)$ on worker productivity. Thus the allocation is ex-post (treating individuals with different h as different types) Pareto-efficient but features insufficient consumption insurance across health types from an ex-ante perspective (relative to the social planner's solution, which implements complete consumption insurance against health status, $c^{SP}(h) = c^{SP}$).

Note that while it follows trivially from our assumptions that the worker's net pay, $w^{CE}(h) - P^{CE}(h)$, is increasing in h, it is not necessarily true that his gross wage, $w^{CE}(h)$, is increasing in h as well, since equilibrium health expenditures are decreasing in health status. In Appendix D we provide a sufficient condition for the gross wage schedule to be monotonically increasing in health status.[19]

Given the results in this subsection it is plausible to expect that, statically, policies that prevent wages $w^{CE}(h)$ and insurance premia $P^{CE}(h)$ from depending on health status will restore the full efficiency of the policy-regulated competitive equilibrium. We show next that this is indeed the case, providing a normative justification for the two policy interventions within the static version of our model.

Competitive Equilibrium with a No Prior Conditions Law ($i = NP$) First, consider government intervention in the health insurance market. The objective of the government is to prevent consumption risk induced by health insurance premium risk, replacing (2) with (4). The next proposition characterizes the resulting regulated equilibrium allocation:

Proposition 7 *The equilibrium health expenditures under a no prior conditions law satisfies, for each $\tilde{h} \in H$, $x^{NP}(\varepsilon, \tilde{h}) = \max[0, \varepsilon - \bar{\varepsilon}^{NP}(\tilde{h})]$, with cutoffs uniquely determined by*

$$-F_2(\tilde{h}, \bar{\varepsilon}^{NP}(\tilde{h})) = \frac{\sum_h u'(w^{NP}(h) - P^{NP})\Phi(h)}{u'(w(\tilde{h}) - P^{NP})}. \tag{9}$$

The equilibrium wage, for each health status h, is given by equation (3), and the health insurance premium is determined by equation (4), evaluated at the no prior conditions expenditure schedule $x^{NP}(\varepsilon, h)$.

The health expenditure levels are no longer efficient for each health type (as they were in the competitive equilibrium) but provide additional partial consumption insurance against initial health status *across* health types by adjusting the cutoff levels $\bar{\varepsilon}^{NP}(h)$, in the absence of direct insurance against health-induced low wages. As shown in the next proposition, it is efficient to over-insure households with *bad* health status and under-insure those with *good* health status, relative to the first-best.

Proposition 8 *Let $\tilde{h}$ be the health status whose marginal utility of consumption is equal to the population average, i.e. for $\tilde{h}$, $u'(w(\tilde{h}) - P) = \sum_h u'(w(h) - P)\Phi(h)$ so that $-F_2(\tilde{h}, \bar{\varepsilon}^{NP}(\tilde{h})) = 1$. Then, $\bar{\varepsilon}^{NP}(h) < \bar{\varepsilon}^{SP}(h)$ for $h < \tilde{h}$; $\bar{\varepsilon}^{NP}(h) = \bar{\varepsilon}^{SP}(h)$ for $h = \tilde{h}$; and $\bar{\varepsilon}^{NP}(h) > \bar{\varepsilon}^{SP}(h)$ for $h > \tilde{h}$. The cutoffs $\bar{\varepsilon}^{NP}(h)$ are strictly monotonically increasing in health status h.*

This feature of the optimal health expenditure with a no prior conditions law also indicates that mandatory participation in the health insurance contract is an important part of government regulation, since, in the allocation described above, healthy households cross-subsidize the unhealthy in terms of insurance premia *and* they are given a less generous health expenditure plan (higher thresholds) than the unhealthy.

Competitive Equilibrium with a No Wage Discrimination Law ($i = NW$) Now we turn to the effects of government regulation on the labor market. The allocative consequences of the law are summarized in the following proposition, whose proof follows from the first-order conditions of program (7).

Proposition 9 *The equilibrium health expenditures under a no wage discrimination law alone satisfies, for each $\tilde{h} \in H$, $x^{NW}(\varepsilon, \tilde{h}) = \max[0, \varepsilon - \bar{\varepsilon}^{NW}(\tilde{h})]$, with cutoffs determined by*

$$-F_2(\tilde{h}, \bar{\varepsilon}^{NW}(\tilde{h})) = \frac{u'(w^{NW} - P^{NW}(\tilde{h}))}{\sum_h u'(w^{NW} - P^{NW}(h))\Phi(h)}. \tag{10}$$

The equilibrium wage is determined by equation (5) and the health insurance premium is given by, for each h, equation (2), evaluated at the health expenditure profile $x^{NW}(\varepsilon, h)$.

Unlike in the no prior conditions case, we cannot establish monotonicity in the cutoffs $\bar{\varepsilon}^{NW}(h)$. Under a no prior conditions law the regulatory authority partially insures consumption of the unhealthy by allocating higher medical expenditure to them. Under a no wage discrimination law instead, there are two opposing

[19]There we also show that the same health expenditure and consumption allocation can be achieved through *separate* wage contracts offered by competitive production firms and health insurance contracts offered by competitive health insurers.

forces. On one hand, a one unit increase in medical expenditure $P(h)$ is more costly to the unhealthy, since the marginal utility of consumption is higher for this group. On the other hand, production efficiency calls for higher medical expenditure for the unhealthy, given our assumption of $F_{12} \geq 0$ (as was the case for the no prior conditions law). Thus the cutoffs $\bar{\varepsilon}^{NW}(h)$ need not be monotone in h.[20]

Competitive Equilibrium with Both Policies ($i = B$) Combining both a no wage discrimination law and a no prior conditions legislation restores the efficiency of equilibrium, since both policies jointly provide full consumption insurance against bad health h, and the assumed efficient regulation of the health insurance market ensures that the health expenditure schedule is efficient as well. This is the content of the next proposition, whose proof follows trivially from the fact that maximizing (7) subject to (4) and (5) is equivalent to the social planner's problem analyzed in Section 4.1.1.

Proposition 10 *The unique competitive equilibrium allocation in the presence of both a no wage discrimination and a no prior conditions law implements the socially efficient allocation in the static model.*

4.1.3 Summary of the Analysis of the Static Model

The competitive equilibrium implements the efficient health expenditure allocation but does not insure households against initial health conditions. A no prior conditions law and a no wage discrimination law provide partial, but not complete, insurance against this risk. The health expenditure schedule is distorted when each policy is implemented in isolation, relative to the social optimum, as the government provides additional partial consumption insurance through health expenditures. Only both laws in conjunction implement a fully efficient health expenditure schedule and full consumption insurance against initial health conditions h, and therefore restore the *static* first-best allocation.

4.2 Analysis of the Dynamic Model

Having characterized consumption allocations within a period, we now turn to the full dynamic model. Since there is no aggregate risk, the sequence of cross-sectional health distributions $\{\Phi_t\}_{t=0}^T$ is deterministic. Furthermore, conditional on Φ_t today, the health distribution tomorrow is completely determined by the effort choice $e_t(h)$ of households (or the social planner), so that the cross-sectional health distribution evolves as:

$$\Phi_{t+1}(h') = \sum_h Q(h'; h, e_t(h))\Phi_t(h). \tag{11}$$

Under each policy, given a sequence of aggregate distributions $\{\Phi_t\}_{t=0}^T$ we can solve the dynamic maximization problem of an individual household for optimal effort decisions $\{e_t(h)_{h\in H}\}_{t=0}^T$. For this, in this section we assume that the continuation utility after retirement is independent of health status (and normalized to zero): for all $h \in H$, $v_{T+1}(h) = 0$. We relax this assumption in our empirical implementation.

A sequence of optimal effort choices in turn implies a new sequence of aggregate distributions via (11). Solving competitive equilibria then amounts to iterating on the sequences $\{\Phi_t, e_t\}$. Within each period, the timing of events follows exactly that of the static problem in the previous section.

4.2.1 Constrained Social Planner's Problem

As a point of comparison for equilibrium allocations (without and with policies), we first again study the solution of a planner choosing constrained-efficient allocations. Statically, the planner can provide full consumption insurance against initial health conditions, as could both policies. In the dynamic model with endogenous effort choice, an *unconstrained* planner in addition could *dictate* effort choices, whereas both policies under consideration can impact effort choice only indirectly, through changing the economic consequences of worse health outcomes. Thus it is more instructive for comparison to study a constrained

[20] The cutoff condition does imply that agents with high marginal utility relative to the average will have lower cutoffs. So long as these agents still correspond to less healthy agents, we can conclude that the cutoffs for the less healthy have been distorted downward, and vice versa for the healthy relative to the social optimum.

planner's problem in which the social planner has to respect the intertemporal optimality condition with respect to household effort choice $\{e_t(h)\}$, given the age- and health-dependent consumption allocation $\{c_t(h)\}$ chosen by the planner. We think of these constraints as emerging from the inability of the planner to observe household effort choices: if a certain effort $e_t(h)$ is desired by the planner, it has to be induced by a consumption allocation that makes providing that effort individually rational, given the health-dependent consumption allocations from tomorrow onward. For comparability with the competitive equilibrium and its static contract, we also restrict the social planner to allocations that only depend on current age and health (t, h). Let $V_t(h)$ denote the expected lifetime utility for a household with current age t and health status h, given recursively by

$$V_t(h) = u(c_t(h)) - q(e_t(h)) + \beta \sum_{h'} Q(h'; h, e_t(h)) V_{t+1}(h'),$$

with exogenous terminal conditions $\{V_{T+1}(h') \equiv 0\}$. The social planner solves

$$\max_{\{c_t(h), e_t(h)\}} V(\Phi_0) = \sum_h V_0(h)\Phi_0(h) \qquad s.t. \tag{12}$$

$$\sum_h c_t(h)\Phi_t(h) \leq \sum_h \left[g(h)F(h,0) + (1-g(h)) \int f(\varepsilon) \left[F(h, \varepsilon - x(\varepsilon,h)) - x(\varepsilon,h) \right] d\varepsilon \right] \Phi_t(h) \tag{13}$$

$$q'(e_t(h)) = \beta \sum_{h'} \frac{\partial Q(h'; h, e_t(h))}{\partial e_t(h)} V_{t+1}(h'), \tag{14}$$

and the general law of motion given in (11). Equation (13) represents the aggregate resource constraint, where $x(\varepsilon, h) = x^{SP}(\varepsilon, h)$ is the efficient health expenditure schedule characterized in section 4.1.1. The second constraint is the incentive constraint on effort. It equates the marginal utility cost of effort today, $e_t(h)$, to the marginal benefit of better health from tomorrow on, *given* a consumption allocation chosen by the planner and encoded in $\{V_{t+1}(h')\}$.

Equation (14) demonstrates the trade-off for the constrained social planner also present in the evaluation of both policies. Statically, it is optimal for the planner to provide full consumption insurance. Although she can certainly implement such an allocation, it would lead to identically zero effort in all periods; see equation (14). Since the marginal cost of providing effort at $e_t(h) = 0$ is zero (assumption 1) and the benefit for health transitions and thus net production and average consumption is positive (on account of assumptions 3 and 4), starting from the full consumption insurance and zero effort allocation, a marginal increase in effort is welfare improving (since the consumption insurance losses are second order when starting at the full consumption insurance allocation). More formally, the constrained efficient allocation is characterized in the following proposition, proved in Appendix B.

Proposition 11 *The constrained efficient allocation $\{c_t(h), e_t(h)\}$ has zero effort in the last period and full consumption insurance in the first period: $e_T(h) = 0$ and $c_0(h) = c_0, \forall h$. Furthermore, assume Q is iid*[21] *and positive: $Q(h'; h, e) = Q(h'; \tilde{h}, e) > 0, \forall (h', h, \tilde{h}, e)$. Then effort is positive in all periods but the last: $e_t(h) > 0, \forall h, t < T$, and has imperfect consumption insurance with consumption $c_t(h)$ strictly increasing in h in all but the first period.*

This result will be in contrast to the outcome under both policies (see Proposition 12 below), which features full consumption insurance and zero effort, and therefore results in an inefficient allocation.

4.2.2 Dynamic Competitive Equilibrium without and with Policy

In the presence of wage and health insurance policies, households of different health types h interact, since the population health distribution Φ_t determines the pooled wage and health insurance premium and the health expenditure cutoff $\bar{\varepsilon}^i(h; \Phi_t)$. The dynamic program in policy regime $i \in \{CE, NW, NP, B\}$ is:

$$v_t^i(h; \Phi) = U^i(h, \Phi) + \max_{e_t(h)} \left\{ -q(e_t(h)) + \beta \sum_{h'} Q(h'; h, e_t(h)) v_{t+1}^i(h', \Phi') \right\} \qquad where \tag{15}$$

[21] The iid assumption on Q is by no means necessary and can be replaced by the assumption that output produced by health type h net of health expenditures and consumption allocated to type h is strictly increasing in h. This assumption is satisfied in our quantitative analysis, but it does involve an assumption on endogenous variables.

$$U^i(h,\Phi) = \max_{x^i(\varepsilon,h,\Phi),w^i(h,\Phi),P^i(h,\Phi)} u(w^i(h,\Phi) - P^i(h,\Phi)) \tag{16}$$

$$P^i(h;\Phi) = (1-g(h))\int_0^{\bar{\varepsilon}} f(\varepsilon)x^i(\varepsilon,h;\Phi)d\varepsilon, i \in \{CE, NW\} \tag{17}$$

$$P^i(\Phi) = \sum_h \left[(1-g(h))\int_0^{\bar{\varepsilon}} f(\varepsilon)x^i(\varepsilon,h;\Phi)d\varepsilon\right]\Phi(h) \text{ if } i \in \{NP, B\} \tag{18}$$

$$w^i(h;\Phi) = g(h)F(h,0) + (1-g(h))\int_0^{\bar{\varepsilon}} f(\varepsilon)F(h,\varepsilon - x^i(\varepsilon,h;\Phi))d\varepsilon \text{ if } \{CE, NP\} \tag{19}$$

$$w^i(\Phi) = \sum_h \left\{g(h)F(h,0) + (1-g(h))\int_0^{\bar{\varepsilon}} f(\varepsilon)F(h,\varepsilon - x^i(\varepsilon,h;\Phi))d\varepsilon\right\}\Phi(h), i \in \{NW, B\} \tag{20}$$

and $x^i(\varepsilon,h;\Phi)$ is the equilibrium health expenditure allocation from section 4.1.2 under policy regime i.

Section 4.1.2 showed that the equilibrium health expenditure allocation was given by the simple cutoff rule $x^i(\varepsilon,h;\Phi) = \max\left[0, \varepsilon - \bar{\varepsilon}^i(h,\Phi)\right]$, with policy-dependent cutoffs in equations (8), (9) and (10). These cutoffs depend on the cross-sectional health distribution Φ through average marginal utilities of the economy. Thus, how $x^i(\varepsilon,h;\Phi)$ and $w^i(h;\Phi), P^i(h;\Phi)$ and consumption $c^i(h,\Phi)$ depend on the cross-sectional health distribution Φ varies across policy regime i. Note that for the household to solve its dynamic programming problem, it only needs to know the sequence of potentially h-contingent wages and health insurance premia $\{w_t^i(h), P_t^i(h)\}$, but not necessarily the sequence of distributions that led to it.[22]

Given such a sequence, the dynamic programming problem of the household then reads as

$$v_t^i(h) = u(w_t^i(h) - P_t^i(h)) + \max_{e_t(h)}\left\{-q(e_t(h)) + \beta\sum_{h'} Q(h';h,e_t(h))v_{t+1}^i(h')\right\} \tag{21}$$

with exogenous terminal condition $\{v_{T+1}(h)\}$. As before the optimality condition reads as

$$q'(e_t(h)) = \beta\sum_{h'}\frac{\partial Q(h';h,e_t(h))}{\partial e_t(h)}v_{t+1}^i(h'). \tag{22}$$

By assumptions 1 and 3, from equation (22) it follows that effort $e_t(h)$ is positive for all t and h as long as $v_{t+1}^i(h')$ is strictly increasing in h'. Note that, although equation (22) looks identical across policy regimes, the determination of the value functions on the right-hand side of the equation is not. Extra consumption insurance induced by policy, ceteris paribus, reduces the variation of v_{t+1} in h' and thus limits the incentives to exert effort in order to achieve a (stochastically) higher health level tomorrow.

Also note that, in a competitive equilibrium, the expenditure cutoff and thus the health spending allocation of a health type h are independent of the cross-sectional health distribution; there is no interaction of health types at all. With one or both policies in place, however, the cross-sectional distribution of health types in society affects individual households through the aggregate wage $w_t = w(\Phi_t)$ and/or the aggregate health insurance premium $P_t = P(\Phi_t)$ as well as through the health expenditure cutoffs. Through both wage and premium pooling and distorting health expenditure allocations, the policies provide partial consumption insurance against health risk, but at the expense of reducing incentives for health effort. In fact, in the absence of direct utility benefits from good health status h (that is, assuming $v_{T+1}(h) \equiv 0$), we have:

Proposition 12 *Suppose there is a no wage discrimination and a no prior conditions law in place simultaneously. Then $e_t^B(h) = 0$ for all h and t. The provision of health insurance is socially efficient. From the initial distribution Φ_0, the health distribution in society evolves according to* (11) *with* $e_t(h) \equiv 0$.

It follows from Propositions 11 and 12 that effort is inefficiently low if both policies are in place, since there are no incentives, through wages or health insurance premia, to exert effort to lead a healthy life. We now map our model to cross-sectional health and effort data from the PSID and the MEPS to quantify the insurance and incentive effects of these government regulations, and deduce their consequences for the evolution of the cross-sectional health distribution, as well as aggregate production, consumption and health expenditures.

[22] Appendix F describes our computational algorithm that exploits this insight.

5 Mapping the Model to the Data

Prior to spelling out the details of the empirical implementation we now discuss what basic aspects of the data will drive the model parameterization and thus the quantitative results. This discussion will also justify the extensions of the model for the empirical analysis.

First, as documented in section 2, the data display a *positive correlation between health and wages.* In the model the impact of health status on wages is determined by the production function $F(h, \varepsilon - x)$. The strength of this dependence on h determines both the scope of the wage insurance benefits against health-induced fluctuations and the efficiency costs of a deteriorating population health distribution.

Second, the data in section 2 display a *negative correlation between health status and health expenditures* that will inform the estimation of the probability $g(h)$ of receiving a negative health shock ε. The strength of this dependence in turn will quantify the insurance benefits and efficiency cost of a no prior conditions legislation limiting the dependence of health insurance premia on health status h.

The *third* key observation of section 2 was that *health effort e raises, probabilistically, future health status.* In the model this effect is encoded in the health transition function Q. The three "technology" functions $F(h, \varepsilon - x)$, $g(h)$, $Q(h'|, h, e)$ thus capture the impact of health status and health efforts on economic outcomes, and the utility and cost functions $u(c)$ and $q(e)$ measure its welfare consequences.

5.1 Extending the Model

To obtain empirically plausible magnitudes of the three key effects discussed above in our model, and achieve a satisfactory model fit to the micro data, we augment the model along five dimensions. First, in the data, the strong association of wages and health spending with health status may be driven either by the direct impact of health or by other observable or unobservable factors correlated with health status, income and health expenditures. Thus, to obtain estimates of the direct effect of health status h, in a first step, we permit wages and health expenditures to also depend on other observable individual characteristics besides health status, such as gender, race, education and age. As appendices G.4 and G.5 explain in detail, we purge a subset of these observables (race and gender) from the data through regression analysis, since they are orthogonal to our model. In contrast, since the model has an explicit age (t) dimension, and since educational ($educ$) differences in wages and health effort are informative about the estimation of the model (as described in the next subsection), we capture these observables explicitly in the model by specifying production as $F(t, educ, h, \varepsilon - x)$, health shock probabilities and its distribution as $g(t, h)$, $f(\varepsilon, t)$ and the health transition functions as $Q(h'|h, e; educ)$. We consider two education groups: individuals who completed high school or less, and individuals who have at least some college education. Moreover, in section 7.1.1 we address the issue of unobserved heterogeneity across health status cells $h_1, ..., h_4$ by a) exploring the properties of the residuals from the regressions controlling for observables, b) documenting the robustness of our results to a smaller wage-health gradient and c) arguing that the impacts of health on wages implied by our model estimates fall within the broad range of estimates from the empirical literature.

Second, in the benchmark model the only purpose of medical expenditures is to offset productivity-reducing health shocks ε. In the data some households have health expenditures from catastrophic illnesses that exceed their labor earnings in a given year. The benchmark model cannot rationalize these very rare but large expenditures. We therefore introduce a second, catastrophic health shock z. When individuals receive this shock, they *have to* spend z; otherwise, they incur a prohibitively large utility cost. We denote the mean of this (age- and education-dependent) shock by $\mu_z(t, educ, h)$. Since this shock will be fully insured, it simply shifts up health insurance premia by $\mu_z(t, educ, h)$ to empirically plausible levels.

Third, in the model health effort e is one dimensional, whereas in the PSID data we observe three effort measures (light exercise, heavy exercise, the intensity of smoking). To accommodate multi-dimensional effort choices, we now assume that overall effort e is a weighted sum of the different types of health efforts.[23]

Fourth, the model thus far assumed that all households have the same disutility from health effort. The data suggest that households with worse health status tend to exercise less; in order to capture this empirical regularity, we augment the model with a health-specific preference shifter $\gamma(h)$ so that the cost of health

[23] In section 7.1.3 we extend the model to allow for other unobserved (and thus mismeasured) components of effort and demonstrate that our policy conclusions are robust to this measurement error in the data.

effort becomes $\gamma(h)q(e)$. Note that since $\gamma(h)$ only affects the disutility of effort, which is separable from the utility of consumption, the static analysis in section 4.1 remains unchanged, and in the dynamic analysis of section 4.2, only the marginal cost of effort changes to $\gamma(h)q'(.)$ in equations (14) and (22).

Finally, thus far we have assumed that the continuation utility at retirement was independent of the health status h, and more broadly, that health status only had consequences for labor productivity, but not for lifetime utility directly. We now introduce an education-dependent terminal direct utility payoff to health $v_{T+1}(educ, h)$ that impacts the incentives to provide health effort late in the household's working life and helps the model to match the health effort profile of older workers. This extension of the model avoids the complete collapse of incentives under both policies (that is, Proposition 12 from the theoretical model no longer applies). It captures the direct utility benefits from better health due to higher quality of life and increased longevity after retirement. Thus, although the main mechanism we stress in this paper works through the productivity benefits of better health in working life, through the terminal value function $v_{T+1}(educ, h)$ we capture the direct health utility benefits or health-dependent mortality that the literature has modelled (see, e.g, Hall and Jones (2007) or Ales et al. (2014)). We now turn to the determination of the parameters of the augmented model.

5.2 Parameter Estimation

We assume that one model period is six years, a compromise between ensuring that effort has a robust effect on health transitions (which requires a sufficiently long time period) and reasonable sample sizes for estimation of a subset of the model parameters (which favors short time periods).[24] We divide the model parameters into three broad sets: a small subset of preference parameters that we choose outside the model, a second set of parameters governing health expenditure and health transitions that we estimate directly from the data, and a third set of parameters governing the production function, health shock probabilities, the health-shock distribution, and the remaining preference parameters that are estimated inside the model to match moments from PSID and MEPS data.

5.2.1 A Priori Chosen Parameters

We fix the values of the coefficient of relative risk aversion σ and time discount factor β a priori. Consistent with values used in the quantitative macroeconomics literature, we choose $\sigma = 2$ and an *annual* $\beta = 0.96$.

5.2.2 Parameters Estimated Directly from the Data

In the second step we estimate the subset of the model parameters governing $Q(h'|h, e; educ)$ and $\mu_z(t, educ, h)$ directly from the data, without having to rely on information from the model equilibrium. We also measure the initial cross-sectional health-education distribution $\Phi_0(h, educ)$ directly from the PSID data.

Health Transition Function $Q(h'|h, e; educ)$ The only choice affecting health status transitions is effort e. With panel data from the PSID on health status that also contain information on measures of e, the transition function Q can be estimated directly from the data. From 1999, the PSID contains measures of light and heavy exercise (e^l, e^h) and cigarettes smoked per day s. We normalize each measure to lie between 0 and 1 and then estimate the following specification of Q, with parameters allowed to be education-specific.[25]

$$Q(h'; h, e) = \begin{cases} \left(1 + [\lambda_1(h)e^{\lambda_2(h)}]^{\zeta_i(h)}\right) G(h, h'), & \text{if } h' = h + i, h > 1 \text{ or } h' = h + 1 + i, h = 1, i \in \{1, 2\} \\ \left(1 + \lambda_1(h)e^{\lambda_2(h)}\right) G(h, h'), & \text{if } h' = h, h > 1 \text{ or } h' = h + 1, h = 1 \\ \left(\dfrac{1 - \sum\limits_{h' \geq h} Q(h'; h, e)}{\sum\limits_{h' < h} G(h, h')} \right) G(h, h'), & \text{if } h' < h, h > 1 \end{cases}$$

[24] The longer period length also makes the assumption that health shocks are *iid* conditional on current h more plausible.

[25] Light exercise is measured as the number of days an individual carries out light physical activity (walking, dancing, gardening, golfing, bowling); heavy exercise is measured symmetrically and includes heavy housework, aerobics, running, swimming, or bicycling. The normalized measure of physical activity is the share of days light and physical activities are performed. For smoking, we use the number of cigarettes smoked per day with 50 cigarettes as maximum, normalized to 1.

The idea behind this specification is that, in the absence of health effort, there is a baseline health transition probability given by $G(h, h')$. This probability can be modified by health effort e, whose effectiveness is governed by the parameter vector $\{\lambda_1(h), \lambda_2(h), \zeta_1(h), \zeta_2(h)\}$ where the $\lambda_i(h)$ capture the effect of effort on maintaining current health status, and the $\zeta_i(h)$ govern the importance on effort for *improving* health status over time. As light and heavy exercise and *not* smoking effort $(1 - s)$ may have different effects on health transitions, we allow for weights on these measures to differ, denoting them as $\{\delta_l, \delta_h, \delta_s\}$, respectively, with $\delta_l + \delta_h + \delta_s = 1$. Our theoretical effort variable corresponds to $e = \delta_l e^l + \delta_h e^h + \delta_s(1 - s)$.

To capture well the full cross-sectional heterogeneity in effort-induced health transitions, we estimate the Q process using a (completely standard) maximum likelihood procedure on individual data described in detail in Appendix G.3. Figure 4 and Table 4 provide a summary of the estimated health transition functions, and their dependence on the observed health effort choices. Each initial health status $h \in \{h_1, .., h_4\}$ accounts for one of four sub-panels of Figure 4 and plots the estimated probabilities as well as the data-implied frequencies[26] of transitioning into health status $h' \in \{h_1, .., h_4\}$ as a function of health effort e. The estimated transitions capture well the fact that, in the data, higher health effort is associated with a larger frequency of favorable health transitions, and for most (h, h') pairs, matches well the gradient with respect to e. The exceptions are concentrated among e groups with few observations (represented by the small size of the markers), which the maximum likelihood procedure using individual data down-weighs in importance.

Table 4: Estimated Parameters for Health Transition Function

Parameters	Low Education	High Education
$G(h, h')$	$\begin{bmatrix} 0.883 & 0.090 & 0.023 & 0.003 \\ 0.739 & 0.233 & 0.027 & 0.002 \\ 0.174 & 0.584 & 0.225 & 0.018 \\ 0.066 & 0.267 & 0.630 & 0.037 \end{bmatrix}$	$\begin{bmatrix} 0.972 & 0.024 & 0.003 & 0.003 \\ 0.743 & 0.245 & 0.012 & 0.001 \\ 0.086 & 0.519 & 0.296 & 0.099 \\ 0.044 & 0.203 & 0.681 & 0.071 \end{bmatrix}$
$\{\delta_l, \delta_h, \delta_s\}$	{ 0.039 0.392 0.568 }	{ 0.050 0.378 0.572 }
$\lambda_1(h), h = \{h_1, ..., h_4\}$	{ 2.443 1.082 1.458 12.397 }	{ 21.294 1.087 1.023 8.534 }
$\lambda_2(h), h = \{h_1, ..., h_4\}$	{ 0.709 0.026 0.244 0.880 }	{ 0.976 0.013 0.736 0.942 }
$\zeta_1(h), h = \{h_1, ..., h_3\}$	{ 2.126 30.457 5.562 }	{ 1.255 39.776 5.322 }
$\zeta_2(h), h = \{h_1, h_2\}$	{ 3.702 49.642 }	{ 1.493 56.684 }

As further justification that the estimated impact of health effort on health transitions is plausible, we relate the implications of our estimates to empirical studies that have estimated the impact of health efforts on the prevalence of *specific health conditions*. Using data from MEPS, summarized in Table 16 in Appendix G.3, we map our measure of health status h into the prevalence of a specific disease such as hypertension or diabetes. We then divide our PSID sample into four health statuses and five effort quintiles and use our estimated health transition function to determine the impact of effort e on the incidence of a specific health condition in the future, as implied by our estimated Q. As detailed in Appendix G.3, our findings suggest that after controlling for initial health status, high effort (highest e quintile) lowers the prevalence of diabetes between 20 to 24% (relative to the lowest effort quintile), whereas in the randomized, controlled trials conducted by Tuomilehto et al. (2001) and Hamman et al. (2006), the prevalence decreases by 44-58%. We conduct similar exercises for other diseases and find that our estimated impacts lie within the findings in the medical literature summarized there.[27] This provides further evidence that the quantitative impact of

[26] The lines are the model-implied estimated transition probabilities. To summarize the data, for each initial h we group all households with that h into five bins according to their observed effort choices and scatter-plot their transition frequencies. The size of each (diamond, circle, star, square) represents the number of observations in the respective bin. Figure 4 contains population-weighted averages of transition probabilities across education groups. The estimated transition functions by education are contained in Appendix G.3.

[27] Colman and Dave (2012) divide their sample from the National Health and Nutrition Examination Survey and its follow-ups into *three* exercise groups and report the impact of exercise on various health conditions, using fixed effects to control for unobserved time-invariant fixed factors. They document that with high levels of exercise the incidence of hypertension decreases between 9-16% relative to the mean, interpreting their results as "plausibly causal." Our estimated Q predicts a decline of between 4-6%. Analyses with BMI, diabetes and heart disease give similar conclusions; see Appendix G.3.

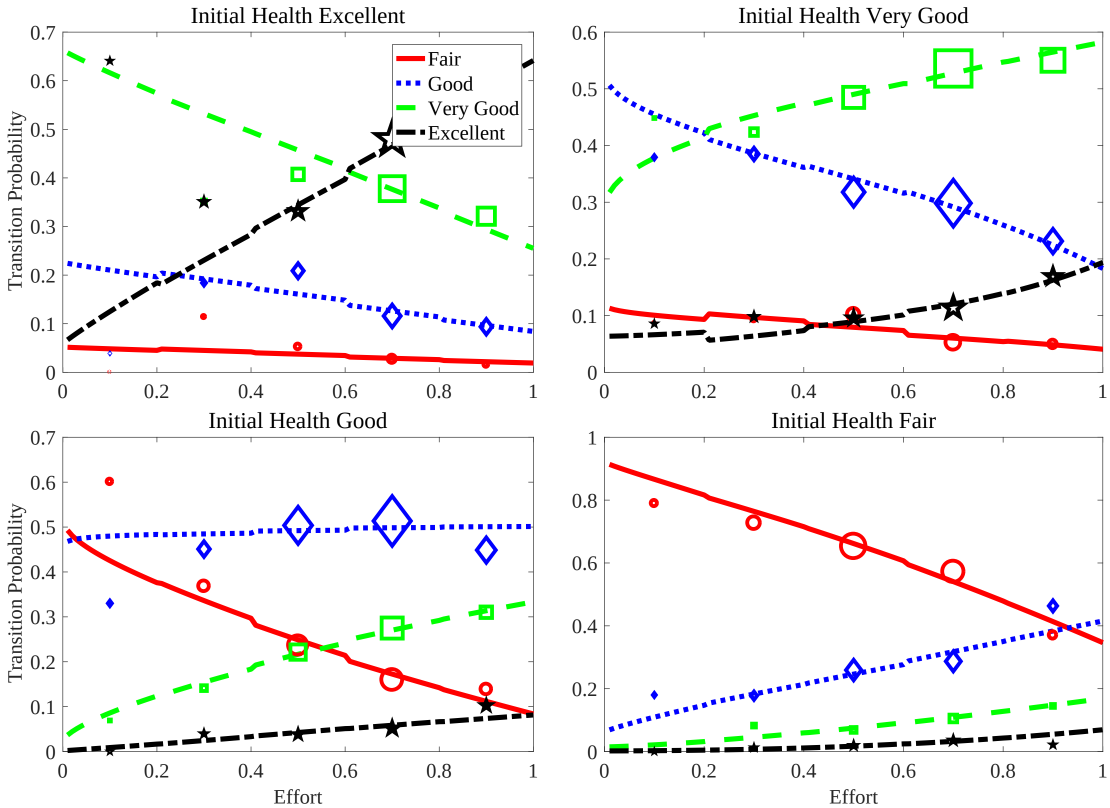


Figure 4: Health Transition Functions - Data (Scatter) vs. Estimated (Line)

health efforts on health transitions is empirically plausible, and in fact, is at the low end of these empirical estimates.

Catastrophic Health Expenditures $\mu_z(t, educ, h)$ In the model, catastrophic health shocks only matter because they shift the age-, education- and health-status-specific expenditure means, and thus the implied health insurance premia. We measure these means directly from MEPS data, defining catastrophic health shocks as those that trigger expenditures larger than \$10,000. Details about the division of health expenditures between catastrophic and "discretionary" (driven by ε) shocks are contained in Appendix G.5.

5.2.3 Parameters Estimated within the Model

We estimate the remaining parameters of the model via the GMM, so that selected model statistics match their empirical counterparts.[28] These parameters govern the production function $F(t, educ, h, \varepsilon - x)$, the health shock probability $g(t, h)$ and the ε shock distribution $f(t, h)$, the curvature of the effort cost function ψ, its health-dependent preference shifters $\gamma(h)$ and the terminal value function $v_{T+1}(educ, h)$. The parameters and the empirical targets used to estimate them are summarized in Table 5.

The structure of our model allows us to estimate the parameters in two separate steps. Since wages and health expenditures are determined exclusively from the *static* part of the model, independent of effort decisions and the associated health evolution in the dynamic part, in a first step we estimate the parameters for the production function and health shock process. In a second step we then employ the *dynamic* part of the model to estimate the remaining parameters.

[28] In contrast to effort choices, which are measured in the PSID as continuous variables and display rich heterogeneity across households, self-reported health status lies on a coarse grid. Thus, when mapping the health-status-related elements of our model into the data, we opt for a GMM approach that matches health-group averages in the data, rather than attempting to exploit the individual variation in wages through maximum likelihood, as we did for the health transition function.

Table 5: Description of Parameters and Moments

Parameters [Number]		Description	Moments [Number]
		Production	
$h_1, ..., h_4$	[4]	Health levels	
$A(t, educ, \tilde{h})$	[28]	Age-educ-health effect	
$\alpha_t(t)$	[7]	Time effect of health	• Smoothed income $w(t, educ, h)$ moments [$7 \times 2 \times 4 = 56$]
$\alpha_e(educ)$	[2]	Education effect of health	
$\phi(educ)$	[2]	Effect of shock, level	• Smoothed medical expenditure $x(t, educ, h)$ moments [$7 \times 2 \times 4 = 56$]
$\xi(educ)$	[2]	Effect of shock, exponent	
		Health Shock	• Fraction with zero medical expenditure by health [4]
$\tilde{g}(h)$	[4]	Prob. of not having a shock	
α_g	[1]	Age effect on probability	• Fraction with zero medical expenditure by age [7]
$\mu_\varepsilon, \sigma_\varepsilon^2$	[2]	Distribution of ε shocks	
$\alpha_\mu, \alpha_\sigma$	[2]	Age effect on distribution	• Mean and variance of medical expenditure [2]
		Effort and Terminal Value	
$\gamma(h)$	[4]	Health-dependent preference	• Average effort by age group, education, and health [$2 \times 2 \times 4 = 16$]
ψ	[1]	Curvature of cost function	
$v_{T+1}(educ, h)$	[8]	Terminal value of health	
Normalizations: $\alpha_e(1) = 1, v_{T+1}(educ, 1) = 0$			
Total number of parameters: 64			Total number of moments: 141

Step 1: Production Function and Distribution of Health Shocks As discussed above, two key ingredients of our quantitative analysis are the functions that determine the impact of health status h on wages and on health insurance premia. In the model, these two are tightly connected as medical expenditure choices impact labor productivity and thus incomes. We capture the large heterogeneity in labor income and medical expenditures across age, education, and health status in the data by a flexible parameterization of the production function and the health shock process in the model, and *jointly* estimate the parameters using the static component of the model.

Concretely, we assume that the production technology takes the following functional form:

$$F(t, educ, h, \varepsilon - x) = A(t, educ, \tilde{h}) h^{\alpha_t(t)\alpha_e(educ)} - A(t, educ, \tilde{h})\phi(educ)\frac{(\varepsilon - x)^{\xi(educ)}}{h} \quad (23)$$

with $\tilde{h} \in \{\{h_1\}, \{h_2, h_3, h_4\}\}$. This specification encodes two impacts of health status on labor productivity. The first term in equation (23) captures the impact of health status, age and education on wages in the *absence of health shocks* and health expenditures ($\varepsilon = x = 0$). The base level of wages $A(t, educ, \tilde{h})$ is determined by flexible age and education effects, and permits a discount for the lowest health group (captured by $\tilde{h}$). The elasticity of wages with respect to health status is permitted to be age- and education-specific, as parameterized by $\alpha_t(t)$ and $\alpha_e(educ)$. The objective of this first component is to capture the mean life-cycle profiles of wages and their dependence on education and health *status* in the data.[29]

The second part of equation (23) encodes the negative productivity impact of health shocks and the offsetting impact of health expenditures. It parameterizes its importance by the education-specific scaling and elasticity factors $\phi(educ)$ and $\xi(educ)$. This term is divided by health status h so that the marginal benefit of health expenditures x declines with better health and thus $-F_{12} < 0$, as assumed in the theoretical model. This functional form, together with the estimated parameters, in part determines how strongly health expenditures x and thus insurance premia increase with lower health status h.

[29]We could have introduced a full set of health-education-age effects, but given the good model fit displayed in Figure 5 we opted for a more parsimonious specification. Note that the health status levels $\{h_1, h_2, h_3, h_4\}$ themselves are also parameters to be estimated, since the categories {Excellent, Very Good, Good, Fair} used in the data have no cardinal interpretation.

The other model component that captures the impact of health status on the health expenditure gradient is the probability $g(t,h)$ of not receiving an adverse health shock ε. We assume $g(t,h) = \tilde{g}(h)\exp(-\alpha_g \times t)$, where $\tilde{g}$ summarizes the positive impact of health status, and α_g represents the negative effect of age on this probability. Conditional on receiving an ε shock, its distribution is determined by the age-dependent probability density $f(\varepsilon; t)$ that we assume to be log-normal.[30] To estimate the age-dependent mean and variance[31] of the ε distributions, we exploit the theoretical result from section 4.1 that medical expenditures x are linear in the shock: $x(\varepsilon, h) = \max[0, \varepsilon - \bar{\varepsilon}(h)]$, and thus the distribution of x coincides with that of the shocks themselves, above the endogenous health-specific threshold $\bar{\varepsilon}(t, educ, h)$. Note that the parameters governing $F(\cdot), g(\cdot), f(\cdot)$ and $\{h_1, h_2, h_3, h_4\}$ have to be estimated jointly, since they combine to determine wages and health expenditures in the static part of the model.

In order to pin down these parameters we use the set of (age-, education- and health-status-contingent) wage and medical expenditure moments summarized in Table 5. As discussed above, the impact of health status on wages (and thus incomes) and on health expenditures is crucial for our quantitative results. The functions $F(\cdot), g(\cdot), f(\cdot)$ must be flexible enough to match the data well; their exact functional forms are secondary. The crucial determinants of the health impact on productivity and expenditures are then the empirical moments we use as targets. In generating these targets, we control for observable factors that may be correlated with health status (such as race and gender), and allow for heterogeneous impacts of health across age groups and education.[32]

As expected, conditioning on observables before deriving the health gradient of incomes and health expenditures lowers the dependence on health status relative to the raw data, but the health effect remains quantitatively very important, as Figure 5 demonstrates.[33] It displays the empirical targets for labor income as well as medical expenditures over the life cycle, and includes the model-implied profiles (Table 22 in the appendix contains the numbers underlying these plots). These match the empirical targets very well, acknowledging that this is due to the flexible parameterization of the production and health shock distribution functions we employ. Figure 5 is the crucial ingredient for the quantitative policy experiments as it summarizes the extent of health-related consumption risk $c(h) = w(h) - P(h)$ and its evolution over the life cycle, in the absence of government social insurance policies.

Conditioning on observables does not rule out that the estimated income and health expenditure profiles might at least partially be driven by an unobserved factor that is correlated with health status, income and/or health expenditures or effort. To investigate whether this form of unobserved heterogeneity explains the joint relation between health status on one hand and income, exercise and health shocks on the other hand, we explore the properties of the regression residuals (after controlling for all observables, including health status) for income, health expenditures and effort. Based on the low cross-sectional correlation of these residuals, we conclude that it is unlikely that an unobservable fixed factor *jointly* drives these variables.[34]

Finally, it is possible that a high individual unobserved fixed factor impacts both an individual's health status and her labor productivity (but not necessarily effort or health expenditures). This could potentially

[30] French and Jones (2004) estimate the cross-sectional distribution of health care costs using HRS (aged 51-61) and AHEAD (aged 70 or older) data and find that a log-normal distribution fits their data well. We could estimate the age-dependent ε distributions non-parametrically from the expenditure data. However, since these data display distributions that are close to log-normal (as French and Jones (2004) found for their data sets), the resulting distributions would be very similar to the age-dependent log-normal ones used here.

[31] The age dependence of means and variances is determined by the four parameters $(\mu_\varepsilon, \alpha_\mu, \sigma^2_\varepsilon, \alpha_\sigma)$ such that $\mu_\varepsilon(t) = \mu_\varepsilon \exp(\alpha_\mu \times t)$ and $\sigma^2_\varepsilon(t) = \sigma^2_\varepsilon \exp(\alpha_\sigma \times t)$.

[32] Concretely, we run a regression of labor income and medical expenditures on race, gender and a full set of age group-education-health dummies and use the regression coefficients on these age group-education-health dummies to construct the empirical education- and health-specific age-income moments. We smooth these moments by fitting them to education- and health-specific quadratic functions of age, and use the predicted profiles as our moments. We select individuals with average labor income over 6 years above \$5,000 to obtain these targets (to focus on working individuals). We conduct the same analysis (whose procedure is contained in Appendix G.4) for health expenditures.

[33] For example, the average ratio of income of those in excellent health relative to fair health falls from 2.2 to 1.7. For medical expenditures, the average ratio of expenditures for fair health relative to excellent health goes down from 4.7 to 3.5.

[34] If an underlying unobserved factor jointly drives these variables at the individual level, the regression residuals should be strongly correlated in the cross-section. Instead, we find that these correlations are quite small. The income and medical expenditure residuals have a correlation of 0.11, the income residual and the effort residual have a correlation of 0.06, and the medical residual and the effort residual correlation have a correlation of 0.02. Also note that the correlations between income and future health, as well as health expenditure residuals and future health are small and negative, suggesting that the causal connection is not running in the reverse direction, from an individual's income and expenditures to future health.

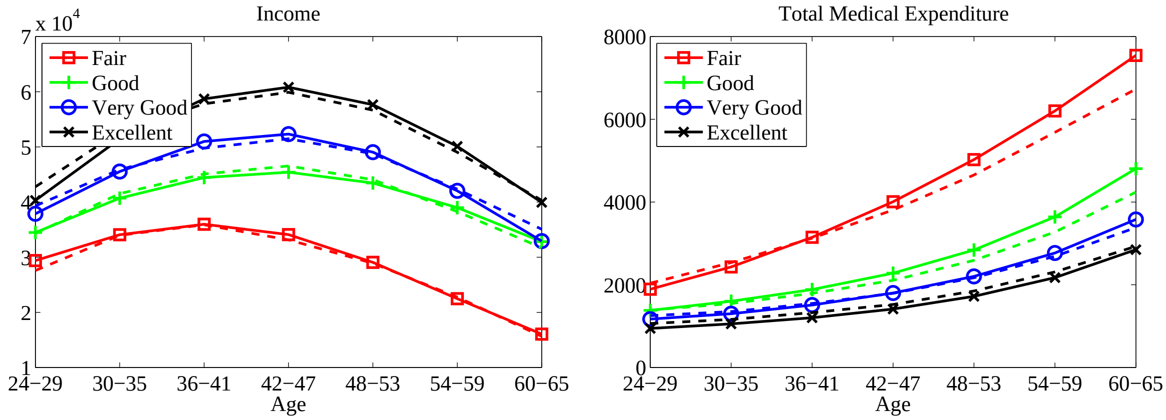


Figure 5: Income and Medical Expenditure - Data (Solid) vs. Model (Dashed)

account for part of the income-health gradient even after controlling for observables. To assess the concern that this specific form of unobserved heterogeneity is responsible for the results of our policy analysis, in section 7.1.1 as a robustness check, we shrink the health-status-related residual income gaps by approximately 60% and demonstrate that the ranking of policies is unaffected by this change (although, expectedly, the welfare gains from the policies become smaller).

As with the effects of health effort, to argue that the income gradient implied by our parameterization is empirically plausible, we now relate them to the empirical literature that has estimated the earnings consequences of work-limiting disabilities. To connect to the literature on disability and earnings (see, e.g., Charles (2003), French (2005), Pelkowski and Berger (2004) and the summary of the literature by O'Donnell et al. (2015)), we map the four-state health status variable h into a two-state grid, reflecting disabled and not disabled workers. In French (2005), workers with bad health comprise 6% (for 30-year-olds) and 30% (for 60-year-olds), and in Pelkowski and Berger (2004), 8% of the population. French (2005) finds wage effects of 8-17%, and hours effects of 20-27% in working-age males. Taking the mid-range effects for wage (12% cut) and hours (23% cut), the implied loss in earnings from bad health is 33%. In Pelkowski and Berger (2004) the income loss is reported to be 52% for workers aged 50 to 62 from the HRS. In our data the percentage of the population with fair health is 7%, which is consistent with the share of workers with bad health in French (2005) and Pelkowski and Berger (2004). The model-based earnings gradient we estimate and use in the policy analysis implies that workers with fair health earn 33% less than the others on average. This reduction drops to 18% in the sensitivity analysis, putting the health gradient we employ within the range of the empirical literature, as represented by French (2005) and Pelkowski and Berger (2004).[35]

Step 2: Health Effort Cost Function and Terminal Value Function With estimates of the production function and health shock distributions in hand, we now use the dynamic part of the model to estimate the remaining preference parameters, using data moments on empirically observed health effort choices (from

[35] An earlier literature estimates the earnings impact of *specific health conditions*. Bartel and Taubman (1979) find significantly negative effects of various health conditions on earnings ranging from 20% to 30%. Mitchell and Butler (1986) estimate a 33% decline in earnings from *arthritis*, when controlling for selection bias, and Mitchell and Burkhauser (1990) estimate an earnings reduction between 19 and 32% from selected health conditions. Normalizing the earnings of workers without disease to one, we can calculate the expected earnings of workers of health status h implied by the empirical estimates of earnings losses due to disease d as $y(d) \times Prob(d|h) + 1 \times (1 - Prob(d|h))$. If we use a 25% income loss for hypertension (the mid-point estimate from Mitchell and Butler) and 33% loss for arthritis, we find that *these conditions alone* lead a loss in expected earnings of 11% to 17% (depending on age), roughly similar to the income ratios used in the robustness analysis of section 7.1.1. Recognizing that the empirical estimates of these income losses only measure the impact of one specific condition, we view these calculations as further confirmation that the health-income gradient implied by our estimates, and the one used in the sensitivity analysis, is well within the range of the empirical estimates. Details of the procedure mapping health status h into specific health conditions and into disability status, as well as an extension of the analysis to hypertension, are included in Appendix G.4.

the PSID). To do so, we assume that the effort utility cost function q takes the functional form

$$\gamma(h)q(e) = \gamma(h)\left[\frac{1}{1-e} - (1+e)\right]^{\psi}.$$

This functional form guarantees that $q''(e) > 0$, that $q(0) = q'(0) = 0$ and that $\lim_{e\to 1} q'(e) = \infty$ as long as $\psi > 0.5$, which is true in our estimation. The parameter ψ controls the elasticity of the utility cost with respect to effort, $\varepsilon_q(e) = q'(e)/(q(e)/e)$. The health-dependent cost shifters $\gamma(h)$ allow us to account better for the *level* of exercise observed in the data. Finally, we introduce terminal education- and health-dependent continuation utilities $v_{T+1}(educ, h)$ to rationalize that even older workers exert health effort in the model. Since only the differences in the continuation values matter for the choice of optimal effort in the last period T, we normalize $v_{T+1}(educ, h_1) = 0$. We estimate the exercise-related parameter values $\{\psi, \gamma(h), v_{T+1}(educ, h)\}$ such that the model reproduces the average effort levels conditional on age group (young[36] and old), education, and health status (see again Table 5).

Table 6: Model Fit

Moments	**Model**	**Data**	**Moments**	**Model**	**Data**
Share, Zero medical expenditure			Average Effort, Young		
Age 24-29	0.147	0.129	High School, Fair	0.572	0.527
Age 30-35	0.118	0.098	High School, Good	0.590	0.595
Age 36-41	0.090	0.087	High School, Very Good	0.640	0.634
Age 42-47	0.070	0.072	High School, Excellent	0.674	0.674
Age 48-53	0.053	0.055	Some College, Fair	0.579	0.685
Age 54-59	0.041	0.043	Some College, Good	0.567	0.662
Age 60-65	0.033	0.037	Some College, Very Good	0.635	0.693
			Some College, Excellent	0.681	0.730
Fair	0.047	0.041	Average Effort, Old		
Good	0.070	0.082	High School, Fair	0.569	0.562
Very Good	0.083	0.073	High School, Good	0.593	0.619
Excellent	0.119	0.098	High School, Very Good	0.640	0.617
			High School, Excellent	0.670	0.657
Mean, Med. Exp. on ε-shock	1,881	1,614	Some College, Fair	0.618	0.590
Variance, Med. Exp. on ε-shock	1,004,459	3,707,439	Some College, Good	0.603	0.610
			Some College, Very Good	0.661	0.660
			Some College, Excellent	0.704	0.697

The data targets and the model fit with respect to health effort and with respect to health expenditures (unless already summarized in Figure 5) are contained in Table 6, and the detailed parameter estimates are reported in Table 21 in Appendix G.6. We estimate a cost elasticity parameter of $\psi = 3.16$. Note that a *large* elasticity of q means that small changes in effort e induce large changes in the cost of providing e, and from equation (22) the behavioral response in e to changes in policy is *smaller* the larger is the ψ estimate.[37]

In terms of the model fit, as with wages and health expenditures, Table 6 shows that our model is parameterized flexibly enough to capture well the variation in health effort across age, education and health groups. Since these are the key ingredients of our model, we are confident that it provides a plausible laboratory for our quantitative policy analysis, to which we turn next.

[36] Young workers are those aged 41 or less. Partitioning the data into finer age groups results in very small cell sizes.

[37] Empirically, the health-wage gradient is steeper for the high education group, and this group displays higher health effort. Intuitively, in our model ψ pins down how strongly effort responds to the benefits of better health, and thus the empirical variation across education groups in effort and health-wage gradients allows us to estimate ψ.

At the median effort $e \approx 0.5$, we have an elasticity of $\varepsilon_q(e) = 6.4$. To put this number into context, in a standard labor supply problem with additively separable cost for hours, the elasticity of the utility cost with respect to hours worked l is $\nu(l) = 1 + 1/\chi$, where χ is the Frisch labor supply elasticity. Evaluating $\nu(l)$ at typical estimates of $\chi \in [0.2, 1]$ (see, e.g., Blundell et al. (2016) for a representative reference), we conclude that our cost of health effort is more elastic than the typical estimates for hours worked, which implies that health effort is *harder* to motivate through economic incentives than regular labor supply along the intensive margin.

6 Results of the Policy Experiments

The social insurance policies we study involve a trade-off between the benefits of providing consumption insurance against bad health from lower wages and higher insurance premia, and the costs from weaker incentives to exert health efforts, resulting in a worse long-run health distribution in the population. In the next two subsections we present the key quantitative indicators measuring this trade-off: first, the insurance benefits in section 6.1, and second, the adverse incentive effects on aggregate production and health in section 6.2. Then, in sections 6.3 and 6.4, we display the welfare consequences of the policy reforms.[38] We start with the pure versions of the policies for which the trade-off between insurance and incentives can be demonstrated most clearly, before turning to partially effective no wage discrimination legislation.

6.1 Insurance Benefits of Policies

In Figure 6 we plot the coefficient of variation of consumption (within education groups) against age for different policies in the model. The combination of both policies, as predicted by the theory in section 4.1, is fully effective in providing perfect consumption insurance: *within-group* consumption dispersion is zero for all periods over the life cycle if both a no prior conditions law and a no wage discrimination law are in place.[39] Also notice from Figure 6 that a wage non-discrimination law alone goes a long way toward providing effective consumption insurance, since the effect of health status on labor income is quantitatively larger than that on health insurance premia. Thus, although a no prior conditions law in isolation reduces within-group consumption dispersion by 10-30%, depending on age, relative to the unregulated equilibrium, as the figure shows, the *remaining* health-induced consumption risk remains very significant. Finally, a comparison with the constrained efficient allocation suggests that a perfect wage pooling policy provides somewhat too much insurance, whereas health premium pooling alone delivers significantly too little, notably at later ages.

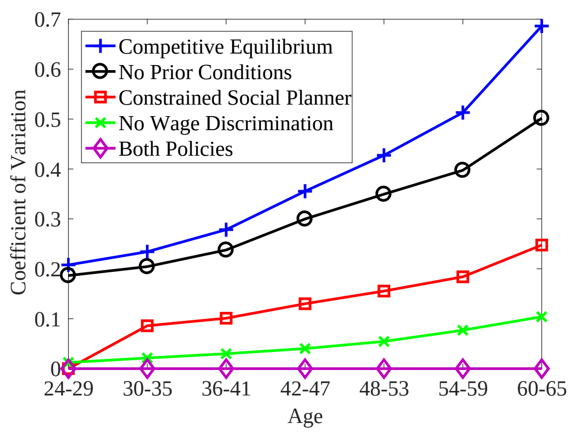


Figure 6: Consumption Dispersion

The consumption insurance described in the previous figure is achieved by implicit transfers across health types: workers do not pay their own competitive actuarially fair price of the health insurance premium or/and they are not fully compensated for their productivity in the presence of the policies. Under the no prior conditions policy, as established in Proposition 8, healthy workers subsidize the premium of the unhealthy. Similarly, wages of the unhealthy workers are subsidized by the healthy, productive workers under the no wage discrimination policy. Figure 7 plots the degree of cross-subsidization over the life cycle, both for households with excellent (left panel) and with fair (right panel) health. The plots for the health insurance premium measure the difference between the actuarially fair health insurance premium a household with a specific health type would have to pay and the actual premium paid in the presence of either a no prior conditions policy or the presence of both policies. Similarly, the wage plots display the difference between the

[38] In the main text we focus on weighted averages of aggregate variables and welfare measures across education (*educ*) types.

[39] Consumption dispersion in the economy at large remains positive due to heterogeneity in consumption by education groups.

productivity of the worker and the wage received under a no wage discrimination policy or in the presence of both policies. Negative numbers imply that the worker is paying a higher premium, or is paid a lower wage than in the unregulated competitive equilibrium, whereas positive numbers imply that a worker is being subsidized in her labor income or health insurance premium. The units on the y-axis are measured in percent of consumption for the specific age-health group in question.

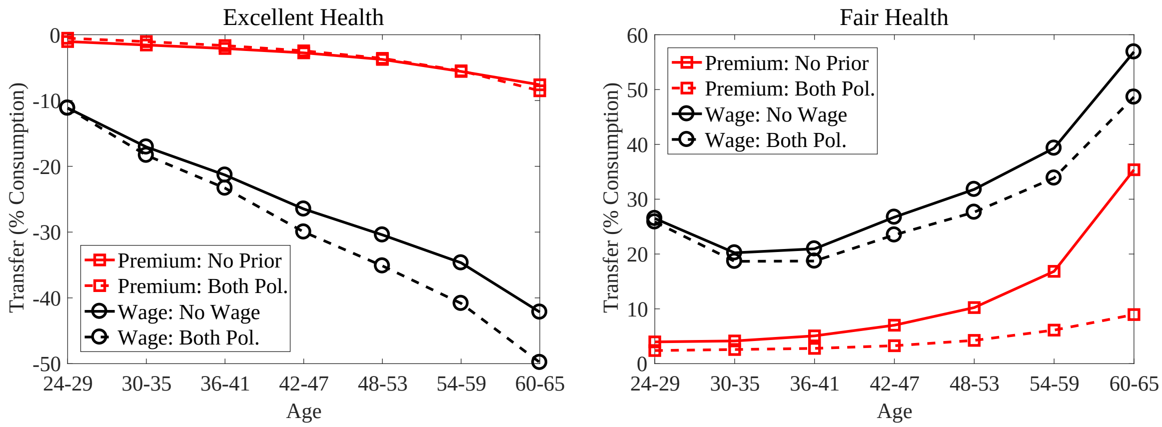


Figure 7: Cross Subsidy: Excellent and Fair Health

From Figure 7 we observe that workers with excellent health significantly cross-subsidize other individuals, both in terms of health insurance premia and in terms of wage transfers. Under the wage non-discrimination law, young workers with excellent health are paid 10% less than their productivity; this number rises to $40\% - 50\%$ near retirement (when the productivity gap between healthy and unhealthy individuals is especially large; see Figure 5). These implicit transfers benefit individuals with bad health. In relation to their consumption, the right panel of Figure 7 shows that these transfers constitute the main source of consumption, especially at older ages (when not only the productivity differentials by across health groups are large, but so is the share of the population with fair health). Figure 7 again demonstrates clearly that transfers induced by a no prior conditions law are significantly smaller than those implied by the wage law, but still a very importance source of insurance, especially for old individuals with poor health.[40]

6.2 Adverse Incentive Effects on Aggregate Production and Health

We now document the aggregate effects of the policies on production and the health distribution. The incentive costs from each policy are inversely proportional to their consumption insurance benefits discussed in the previous subsection, as Figures 8 and 9 show. In these figures we plot the evolution of average health effort and the share of the population with excellent and very good health, under the various policy scenarios. Effort is highest in the unregulated equilibrium, positive under all policies, but substantially lower in the presence of the non-discrimination laws.[41] The policies that provide the most significant consumption insurance benefits also lead to the most significant reductions in incentives to lead a healthy life. It is the very dispersion of consumption due to health differences stemming from health-dependent wages and insurance premia that induces workers to provide effort in the first place, and thus the policies that reduce

[40] Note that the level of subsidization implied by a given policy depends somewhat on the presence of the other policy. In the presence of both policies the health distribution deteriorates more quickly, and thus implicit wage transfers from individuals of excellent health are larger, yet transfers received by households of fair health are smaller (especially in old age) under both policies than in the presence of only a wage non-discrimination law. In addition, as demonstrated in section 4.1 with only a no prior conditions law in effect, the resulting health insurance contract is more generous for the unhealthy, in order to provide indirect income insurance, an effect absent if both laws are in place. This effect is quantitatively very important, as Figure 11 will demonstrate. Thus the cross-subsidies in insurance premia are quantitatively larger (especially for old individuals) if only a no prior conditions law is present (35% of consumption vs. 10% of consumption for the oldest age group), as the right panel of Figure 7 shows.

[41] For all policies and in all periods the terminal direct utility of health induces positive effort through the continuation values in the dynamic programming problem.

this consumption dispersion the most come with the sharpest reduction in incentives.[42] Whereas a no prior conditions law alone leads to only a modest reduction of effort, with a wage non-discrimination law in place the amount of exercise shrinks more significantly. Finally, if both policies are implemented simultaneously, the *only* benefit from exercise is a better distribution of the post-retirement continuation utility, and thus, effort plummets strongly, relative to the competitive equilibrium.[43]

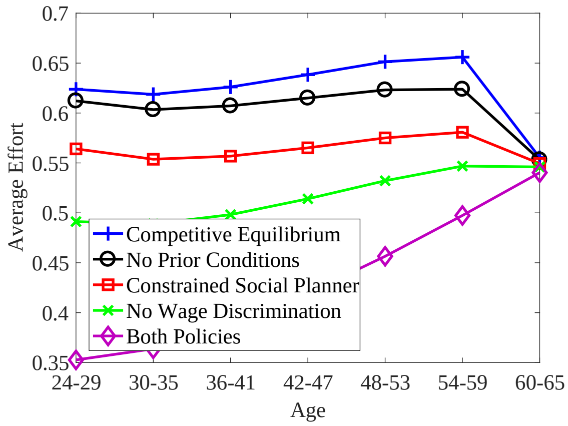


Figure 8: Effort

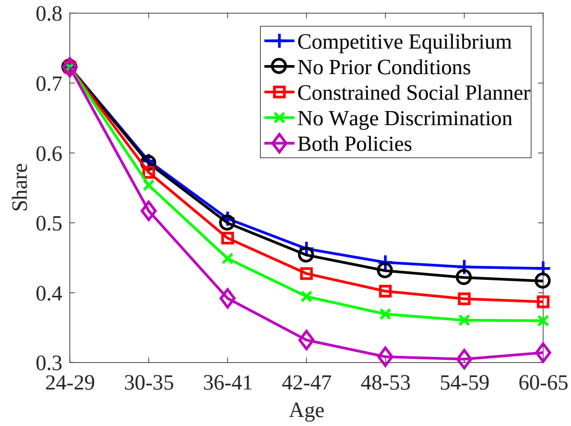


Figure 9: Share with Excellent, Very Good Health

Given the dynamics of effort over the life cycle (and a *policy invariant* initial health distribution), the evolution of the health distribution is exclusively determined by the estimated health transition function $Q(h'; h, e)$. Figure 9 displays the share of households in the population with excellent and very good health and is a direct consequence of the effort dynamics from Figure 8. It shows that health in the population deteriorates under all policies as a cohort ages, but more rapidly if a no prior conditions law and especially if a wage non-discrimination law is in place. The combination of both policies has the most severe impact on public health: under this policy scenario close to 70% enter retirement in less than very good (i.e., fair or good) health, relative to 55% in the competitive equilibrium.

To understand and interpret the large magnitude of the policy-induced decline in the health distribution, recall that we argued in section 5.2.2 that the response of future health to a given change in effort encoded in the estimates of the transition function Q is moderate, and recall from section 5.2.3 that the large estimated elasticity of the effort cost function implies that it is more difficult for economic incentives to move health effort than regular labor supply. The reason effort nevertheless responds so strongly, at least in policy reforms that involve a wage non-discrimination law, is that, in its pure form, this policy is a very large reform, for example, shrinking the dispersion of consumption (as measured by the coefficient of variation) from 0.7 to 0.1 for the oldest age in the model. As Figure 8 suggests, more moderate reforms (such as the no prior conditions law) have fairly small impacts on average effort. Motivated by this observation, in section 6.4 we consider a more moderate wage non-discrimination legislation and document that partial insurance is preferred to the pure form of the law, precisely because of its more benign (and perhaps more realistic) consequences for effort and the health distribution. Finally, note from Figure 9 that even under the most extreme policy reforms, a significant deterioration in the health distribution occurs with substantial delay, in the late 30's of the cohort impacted by the reform already in their mid-20's.[44]

[42]This also explains why average effort is lower in the constrained-efficient allocation relative to the equilibrium allocation.

[43]The impact of the policies on effort is most significant at young and middle ages, whereas, toward retirement, effort levels under all policies converge. This is because the direct utility benefits from better health materialize at retirement and are independent of the non-discrimination laws (but heavily discounted by young households), whereas the productivity and health insurance premium costs materialize throughout the household's entire working life and are strongly affected by the policies.

[44]To place the model-predicted changes in health into historical context we computed, using the mapping between health status h and the incidence of diabetes already used in section 5.2.2, the model-predicted policy induced changes in the incidence of diabetes over the life cycle. Relative to the competitive equilibrium, the model implies an increased incidence of at most 1.5 percentage points under the combination of both fully effective policies, and after a substantial time (starting from age 42 and older). Considering each policy in isolation implies smaller increases of at most 1 percentage point. In the data, from 1996 to 2015 the percentage of individuals between the ages of 45-64 with diagnosed diabetes instead rose by more than 6 percentage

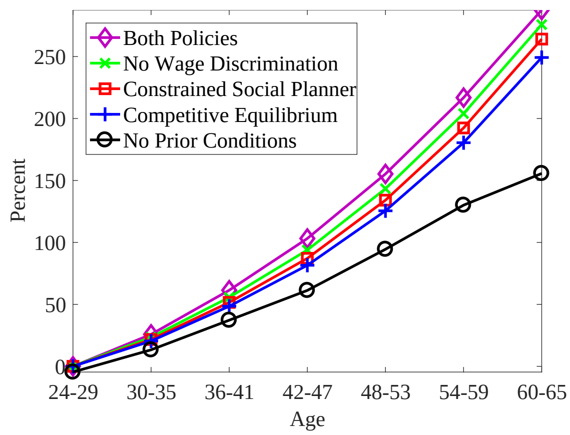


Figure 10: Health Spending

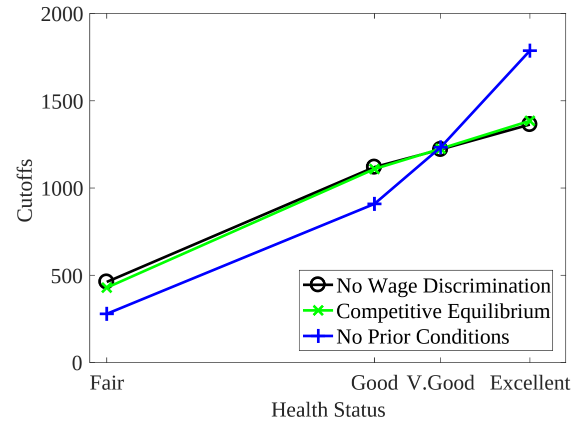


Figure 11: Health Expenditure Cutoffs

The deterioration of the population health distribution has a significant impact on aggregate health spending, output and consumption, as Figures 10, 12 and 13 demonstrate. For a better comparison across policies, the plots are displayed as percent deviations from the value for 24-29 year olds in the competitive equilibrium. Figure 10 shows that the decline of health levels over the life cycle induce higher expenditures on health later in life under all policies, but that the speed of this increase differs strongly across policies.

Health expenditures are determined by two factors: a) the population health distribution (which evolves differently under alternative policy scenarios) and b) the equilibrium health insurance and expenditure contracts, which are fully characterized by the thresholds $\bar{\varepsilon}(h)$ from the static part of the model and that vary across policies. We display these thresholds[45] by health status h in Figure 11. Recall from section 4.1 that the thresholds $\bar{\varepsilon}(h)$ in the competitive equilibrium and in the presence of both policies are socially efficient and thus all three graphs coincide. Also observe that, relative to the efficient expenditure allocation, under the no prior conditions law workers with low health are slightly over-insured (they have *lower* thresholds, $\bar{\varepsilon}^{NP}(h_i) < \bar{\varepsilon}^{SP}(h_i)$ for $i = 1, 2$) and workers with excellent health are strongly under-insured. This is the content of Proposition 8, and it is quantitatively responsible (jointly with relatively high effort and thus a more slowly declining fraction of individuals with excellent health in the population) for the finding that health expenditures are smallest under this policy. The reverse is true under a no wage discrimination law: low health types are under-insured and high types are over-insured, but the differences in the thresholds are minor, relative to the competitive equilibrium. The larger health expenditures later in life in Figure 10 stem primarily from a faster decline in average health status in the population induced by lower health effort.

Finally, Figures 12 and 13 display aggregate production and consumption over the life cycle. Since the productivity of each worker depends on her health and the non-treated fraction of her health shock, aggregate output is lower under policies that lead to a worse health distribution and that leave a larger share of health shocks ε untreated. From Figure 12 we see that the deterioration of health under a policy environment that includes a wage non-discrimination policy is especially severe, fully in line with the findings from Figure 9. The aggregate consumption consequences largely mirror those of output, as Figure 13 shows. Relative to the unregulated equilibrium, a wage non-discrimination law, especially when coupled with a no prior conditions legislation, entails a significant loss of average consumption in society. Note that health expenditures are significantly lower under a no prior conditions law than in the other policy configurations, and therefore, a larger share of output is available for consumption. Consequently, aggregate consumption is actually larger than in the competitive equilibrium early in the life cycle, and not significantly lower in old age.

Overall, the effect on aggregate consumption suggests a quantitatively important trade-off between insurance and incentives. The competitive equilibrium provides strong incentives at the expense of risky consumption, whereas a combination of both policies delivers full insurance at the expense of a significant

points. Thus, by recent historical standards the change in health predicted by the model policy reforms is not implausibly large, although clearly in the data many other factors besides a change in economic incentives might be responsible for this change.

[45]We plot the thresholds for the youngest age-low education group; they look similar for other age-education groups.

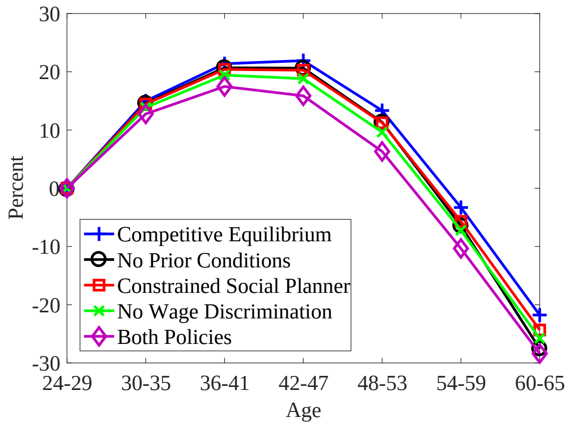


Figure 12: Production

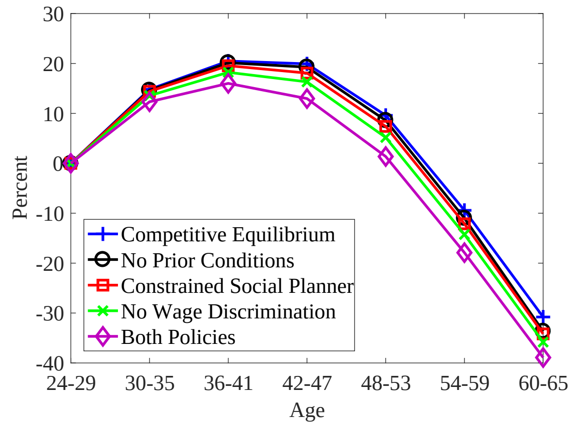


Figure 13: Consumption

deterioration in the population health distribution. The effects of the no prior conditions law on both insurance and incentives are modest, relative to the competitive equilibrium, whereas a no wage discrimination law or both policies insure away most of the consumption risk, but significantly reduce the incentives to exert effort to lead a healthy life, especially early in the life cycle. A comparison with the constrained efficient allocation suggests that the wage policy resolves this trade-off more successfully, but by *fully* insuring the very significant income risks, it leads to a reduction in health incentives that is too large. We now confirm this in the welfare analysis of these hypothetical reforms. In section 6.4 we further strengthen this conclusion by showing that a policy of only partial, but very substantial wage compression fares even better.

6.3 Welfare Implications

6.3.1 Aggregate Welfare

For a fixed initial distribution $\Phi_0(h)$ over health status,[46] denote by $W(c,e)$ the expected lifetime utility of a cohort member, with expectations taken prior to the initial draw h of health, from an arbitrary allocation of consumption *and effort* over the life cycle.[47] Our consumption-equivalent welfare of policy reform $i \in \{SP, NP, NW, B\}$ is given by

$$W(c^{CE}(1+CEV^i), e^{CE}) = W(c^i, e^i).$$

Thus CEV^i is the percentage reduction of competitive equilibrium *consumption* required to make individuals ex-ante indifferent between the competitive equilibrium allocation[48] and that arising under regime i. We also report the welfare implications $SCEV^i$ of the same policy reforms from the static component of the model in section 4.1, in period 0 and again taking as given the initial distribution Φ_0. Thus, $SCEV^i$ provides a clean measure of the static gains from better consumption insurance from the policies against which the dynamic adverse incentive effects have to be traded off.

Table 7: Aggregate Welfare Comparisons

	$SCEV^i$	Dynamic CEV^i
Constrained Social Planner	1.257	5.587
Competitive Equilibrium	0.000	0.000
No Prior Conditions Law	0.192	2.904
No Wage Discrimination Law	1.252	5.055
Both Policies	1.257	2.973

[46] We carry out our analysis for each (*educ*) type separately and report averages across these types.

[47] Using the notation from section 4.2 $W(c^{SP}, e^{SP}) = V(\Phi_0)$ and $W(c^i, e^i) = \int v_0^i(h) d\Phi_0$.

[48] Even the constrained planner problem is solved for each (*educ*) group separately and does not permit ex-ante insurance against being part of a low (*educ*) type. This problem allows a more informative comparison with the competitive equilibrium.

The static welfare consequences reported in the first column of Table 7 are consistent with the consumption dispersion measures displayed in Figure 6. Perfect consumption insurance, as implemented in the solution to the social planner's problem and achieved if both policies are implemented jointly, is worth about 1.3% of competitive equilibrium consumption. The wage non-discrimination law realizes most of these gains, whereas the no prior conditions legislation is significantly less effective in this respect.

The dynamic welfare consequences in column 2 of Table 7 paint a different picture. As in the static analysis, both policies improve on the laissez-faire equilibrium, and the welfare gains are substantial, ranging from 2.9% to 5.1% of *lifetime* consumption. The sources of these welfare gains are improved consumption insurance (as in the static model) and reduced effort (which bears utility costs), which outweigh the reduction in average consumption these policies entail (recall Figure 13). Furthermore, as in the static model a wage non-discrimination law dominates a no prior conditions law because of better consumption insurance.[49]

The key observation of Table 7 is the distinction between the static and the dynamic analysis. Dynamically, it is *not* optimal to introduce a no prior conditions law once a wage non-discrimination law is already in place. The latter policy provides fairly effective consumption insurance, and the further reduction of incentives with the associated fall in mean consumption makes a combination of both policies suboptimal. The welfare losses of pushing social insurance too far amount to about 2% of lifetime consumption. Finally, we see that in contrast to the static case, the best policy (a wage non-discrimination law alone) leads to welfare losses relative to the constrained efficient allocation, although these losses are fairly modest at 0.53% of permanent consumption. They emerge due to inefficiently low consumption insurance, an inefficient effort allocation and an inefficient health expenditure allocation (see again Figure 11), although the latter two effects are quantitatively small. The last effect, however, *is* quantitatively crucial in explaining why the no prior conditions law in isolation fares worse than the wage non-discrimination policy (and a combination of both policies, which restores efficiency in health expenditures; recall Proposition 10).

6.3.2 Heterogeneity by Health Status

The welfare consequences reported in Table 7 were measured before workers learn their initial health level h. They mask very substantial heterogeneity in workers' attitudes toward these policies once their initial health status in period 0 has been revealed. Given the transfers across health types displayed in Figure 7 and the persistence of health status, this is hardly surprising. Table 8 quantifies this heterogeneity by reporting welfare measures computed *after* the initial health status has materialized.

Table 8: Welfare Comparison in the Dynamic Economy Conditional on Health, Age 24-29

	Fair	Good	Very Good	Excellent
Constrained Social Planner	21.403	10.351	5.552	0.119
Competitive Equilibrium	0.000	0.000	0.000	0.000
No Prior Conditions Law	4.989	3.515	2.952	2.041
No Wage Discrimination Law	22.147	10.574	5.053	-1.053
Both Policies	21.448	8.798	2.942	-3.421

Broadly speaking, the lower a worker's initial health status, the more strongly she favors policies providing consumption insurance. For households with very good, good and fair health, the ranking of policies coincides with that in the second column of Table 7, although individuals born with fair health who value insurance a lot are close to indifferent between the ex-ante optimal policy (the wage non-discrimination policy) and having both policies in place. In contrast, young households with *excellent* health prefer the lesser effort distortion (and lesser insurance) from a no prior conditions law to a no wage discrimination law. Note that even for these households, some future insurance through the no prior conditions law is preferred to the competitive equilibrium, even though that policy comes with approximately 1.5% lower consumption, relative to the competitive equilibrium. The differences in the preference for different policy scenarios across different h-households are quantitatively *very* large: whereas fair-health types would be willing to pay 22%

[49]In addition, the no wage discrimination policy induces lower costly effort (Figure 8) but also leads to lower average consumption (Figure 13) compared to the no prior conditions law. These last two effects roughly cancel out in terms of welfare.

of lifetime consumption to see the optimal policy introduced, households with excellent health would be prepared to *pay* 1% of lifetime consumption to prevent precisely this policy innovation.

Table 9: Welfare Comparison in the Dynamic Economy Conditional on Health, Age 54-59

	Fair	Good	Very Good	Excellent
Constrained Social Planner	70.689	14.844	-5.034	-14.682
Competitive Equilibrium	0.000	0.000	0.000	0.000
No Prior Conditions Law	29.619	10.210	-1.252	-5.257
No Wage Discrimination Law	85.272	12.850	-10.134	-22.654
Both Policies	86.645	6.881	-16.896	-29.679

Interestingly, as a cohort ages the assessment of the desirability of the policies under consideration changes. In Table 9 we display the dynamic consumption-equivalent variation measures, but now computed for the second oldest age cohort (aged 54-59). We observe that now policy preferences become more polarized: households with *fair* health now favor *both* policies, whereas those older households with *very good and excellent* health oppose any policy intervention. Given the persistence of health status, an individual in her mid-50's with very good or excellent health will likely spend the remainder of her working life in favorable health and thus does not value insurance against health deteriorations much, relative to the mean consumption losses implied by the policies. Thus, as a cohort ages, the opposition to far-reaching social insurance from those who have maintained very good or excellent health grows stronger.

6.4 Limited Wage Insurance and Resource Cost of Enforcement

In this section, we investigate the impact of no wage discrimination policies such as the ADAAA under the perhaps more empirically realistic assumption that it eliminates some, but not all health related income variation, and that its implementation requires a resource cost. We parameterize the resource cost of the policy by γ and its effectiveness by τ. Labor income of an individual is now determined by the convex combination (with weight $1-\tau$) of individual, health-contingent wages and (with weight τ) average wages:

$$w(h;\tau) = (1-\tau)w(h) + \tau(1-\gamma)w.$$

The perfectly costless full wage-insurance benchmark policy analyzed so far is nested as case ($\tau = 1, \gamma = 0$). As before, consumption is given by the difference between the wage and the health insurance premium.[50] For a given resource cost γ, we think of variations in the insurance parameter τ as either reflecting changes over time in the degree of social insurance afforded by U.S. legislation as the ADAAA augmented the ADA, or as heterogeneity in the extent of disability protection across countries. Alternatively, we can think of τ as a policy parameter chosen by the government, in which case the analysis to follow can be interpreted as solving for the *optimal* degree of social insurance within the restricted set of static, linear insurance mechanisms.

In Figures 14 and 15 we plot, against the degree of insurance τ, aggregate welfare in the absence of a resource cost $\gamma = 0\%$ and with a resource cost of $\gamma = 4.55\%$, the midpoint of Acemoglu and Angrist's (2001) estimates.[51] The benchmark results are subsumed in Figure 14, showing the strong policy preference for a wage non-discrimination law relative to a no prior conditions law, a combination of both policies (the dashed line at $\tau = 1$) or the unregulated equilibrium (partial no wage with $\tau = 0$). Absent a resource cost, Figure 14 shows that as the effectiveness of the policy increases and consumption dispersion declines (τ increases along the x-axis), welfare initially *rises* as better consumption insurance outweighs the effects of reduced effort incentives. However, as wage pooling becomes too effective (for $\tau > 80\%$), the impact of better insurance on effort and the population health distribution becomes too severe, and welfare declines. Figure 14 shows that it is optimal to insure 80% of health-induced income variation (70% if a no prior conditions legislation is also in place). With this degree of insurance, the policy-induced decline in effort is roughly cut in half, and the deterioration of the health distribution is more benign (for details, see Figure 19 in Appendix H.2). Finally,

[50] For each (γ, τ) combination the health expenditure thresholds under the three policies are determined as in section 4.1.

[51] Acemoglu and Angrist (2001) estimate weekly costs of the original ADA of \$24.50 to \$35.00. The costs include lawsuits and accommodations. Relating these costs to average income in our model, these work out to 3.7% to 5.4%.

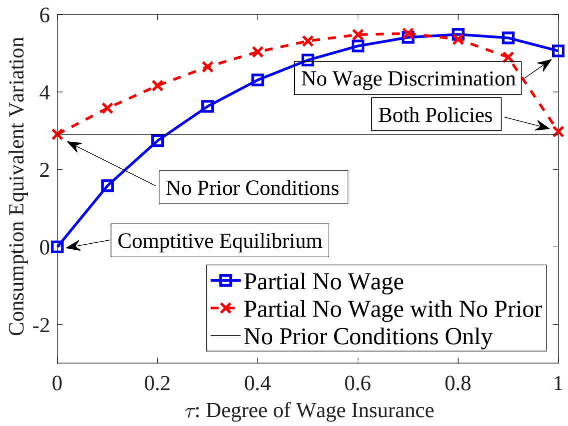


Figure 14: No Resource Cost ($\gamma = 0$)

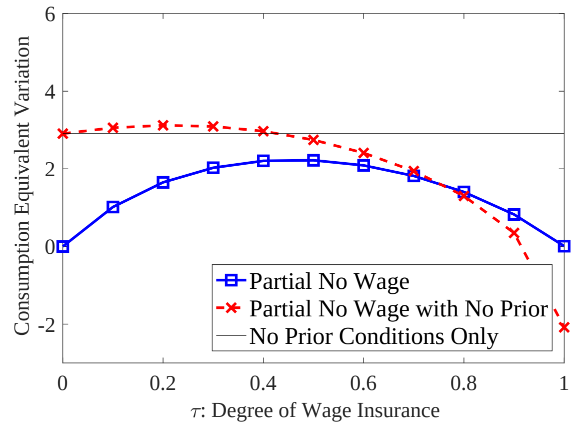


Figure 15: Positive Resource Cost ($\gamma = 4.55\%$)

Figure 15 shows that with costly implementation of the wage non-discrimination law, premium pooling in the health insurance market dominates costly wage pooling. Furthermore, introducing both policies jointly is only optimal if wage insurance is so imperfect that strong incentives for health efforts are maintained.[52]

The results in this subsection confirm the main point of the paper. Empirically, there are significant income and health expenditure risks associated with bad health. At the same time, individuals have some control over their health transitions. Implementing social insurance against health risk is therefore valuable but it comes at an incentive cost. The optimal amount of consumption insurance is very sizable but partial, with non-trivial remaining risk to maintain incentives. With idealized versions of our policies, this leads to the wage non-discrimination law being optimal. With only partial wage insurance which policy combination is optimal depends on the magnitudes of τ and γ. Empirically plausible implementation costs, for example, make the no prior conditions law the preferred policy. The key is that a desirable policy provides strong consumption insurance but leaves partial incentives in place by abstaining from full consumption insurance.

7 Robustness and Sensitivity Analyses

We now explore the robustness of our results to alternative strategies of mapping the model to the data and to model extensions. In subsection 7.1 we investigate how sensitive our results are to the magnitudes of the three key model building blocks. In subsection 7.2 we study how mismeasured *real* health expenditures and a misspecified link between health expenditures and health outcomes would affect our results, and finally, in subsection 7.3 we discuss modelling uninsured households. Table 10 summarizes the results of these exercises.

7.1 Sensitivity Analysis I: Quantitative Importance of the 3 Key Building Blocks

7.1.1 The Link Between Health Status and Income

The first key ingredient of our theory is the impact of health status h on individual income. If this link is weaker than we have estimated, or if households can adjust labor supply to health-induced productivity fluctuations, we might have overestimated the insurance benefits and incentive costs from social insurance. To address the first part of the concern that, even treating income as exogenous, the causal link between health status and labor income (which controlled for observable heterogeneity) might be weaker than estimated, we now conduct a sensitivity analysis with respect to the health effect on income. Concretely, we compress the variation in income profiles across health status by half, while keeping average income in each age group the same.[53] Under the new estimation, workers with excellent health now earn about 29% more than

[52] The total cost is $\gamma\tau w$ and thus increases with τ. With a fixed cost of NW, NP would be preferred to both policies, $\forall\tau$.

[53] Figure 20 in Appendix I.1 displays the old and the new income profiles used as targets for the estimation of the model. In order to keep average income in the economy similar to its level in the benchmark, we scale down the income ratio with respect

Table 10: Dynamic Welfare Results from Sensitivity Analyses

	BM	IG	EL	PL	QM	FX	UN
Constrained Social Planner	5.587	2.218	1.086	2.893	4.115	5.587	5.736
Competitive Equilibrium	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No Prior Conditions	2.904	0.710	1.548	0.564	2.170	1.920	3.071
No Wage Discrimination	5.055	1.045	1.971	2.718	3.697	5.010	5.218
Both Policies	2.973	0.066	0.624	0.945	1.734	2.973	3.140

Note: BM: Benchmark; IG (7.1.1): Lower Effect of Health on Income; EL (7.1.1): Endog. Labor Supply; PL (7.1.2): Insurance Pooling as Benchmark; QM (7.1.3): Re-estimated Q with Mismeasured Effort; FX (7.2.2): Fixed Medical Expenditure; UN (7.3): Model with Uninsured. In PL, the CEV's are relative to the Competitive Equilibrium with insurance pooling.

those with fair health (19% for workers aged 41 or less).[54] As displayed in Table 10 (column labeled IG), since the health-induced income risk is significantly lower, the overall welfare benefits of social insurance policies decline, but remain significant at 1% of lifetime consumption. Crucially, our policy rankings and the conclusion that applying both policies in conjunction is suboptimal is robust to this change.

Second, we extend our model to allow for endogenous labor supply and examine the impact on our results. In Appendix I.2 we use a utility function from Greenwood et al. (1988) that abstracts from income effects to study the endogenous response of labor supply to smooth health-related productivity fluctuations. We show that this leads to a static payoff function of the form:

$$U^{CE}(h) = u\left(\frac{y(h)}{1+\chi} - P(h)\right)$$

where $y(h) = w(h)l(h)$ is labor income, the product between the exogenous wage $w(h)$ and endogenous hours worked $l(h)$, and $\chi \geq 0$ is the Frisch labor supply elasticity. The benchmark model is nested as $\chi = 0$. Thus the importance of a given income gradient in utility is scaled down by the factor $1+\chi$, since individuals now can choose how to smooth utility across health h status by optimal consumption-leisure (labor) decisions. Under the assumption that the health insurance contracts under all policy regimes do not endogenously respond to the introduction of an endogenous labor supply, the scaling of the income vector by $1+\chi$ is the only impact of endogenous labor supply on our dynamic analysis.[55]

To quantify these effects, we retain the calibration of the benchmark model, but use the moderated health-wage gradient from the first part of this section, scale the disutility of labor so that average hours worked in the competitive equilibrium equal 1, choose a Frisch elasticity of $\chi = 0.4$ and redo our analysis. The results are summarized in column EL of Table 10, and reveal that the policy ranking is identical to the first thought experiment (scaling down the wage gradient, IG), but with welfare gains from the policies that lie in between the benchmark model (since individuals now have one more endogenous margin of adjustment) and the first sensitivity analysis (since, relative to column IG, now the income gradient is magnified by an endogenous labor supply, leading to larger health-induced *income* risk).

7.1.2 Underestimating Health Insurance Price Pooling in the Competitive Equilibrium

In our benchmark model individuals also have an incentive to exert health effort because it reduces expected future health expenditures. We thus far have assumed that health insurance premia paid by individuals reflect expenditures conditional on health status, without any pooling. That is, we estimated the model's

to good health $w(h)/w(h=2)$ by half while keeping the income profile of good health the same (in levels).

[54] An alternative interpretation of this analysis is that the original health-income gradient is estimated correctly, but that additional unmodeled government insurance policies or household smoothing mechanisms (e.g., precautionary savings against health shocks, endogenous labor supply, or inter-household transfers) reduce the dependence of consumption on health status h, and by shrinking the health-income gradient we capture these mechanisms, although in a fairly reduced-form way.

[55] The optimal health insurance contract *does* respond to an endogenous labor supply. GHH preferences emphasize the substitution effect, and individuals with better health and higher wages work longer hours in the competitive equilibrium. Thus productivity-enhancing health resources are more beneficial for high-health high-hours individuals. Therefore, the endogenous income gradient $y(h)$ varies *more* by health status than with exogenous labor supply, counteracting the scaling factor $1+\chi$.

competitive equilibrium. It is plausible to argue that the PSID data we used (from 1997 to 2003) were generated by a model in which health insurance premia were already substantially pooled across health types, especially within companies offering group health insurance. Thus, estimating the competitive equilibrium of the model might overstate the health insurance premium risk that households face. We now re-estimate the model under the assumption that all health insurance premium risk is already pooled to address this concern. We refer to this as the insurance-pooling equilibrium.[56]

The re-estimated model needs to match the effort data, with smaller benefits from being healthy. Consequently, although most parameters remain fairly unchanged, the estimated cost shifters $\gamma(h)$ fall and the $q(.)$ function becomes slightly more convex; see Table 24 in the appendix. Crucially, as Table 10 (third column labeled PL) shows, the policy conclusions remain unchanged, but the welfare gains from implementing social insurance policies somewhat diminish. This is to be expected, since the insurance premium pooling in the competitive equilibrium already provides a significant degree of consumption insurance against health risk.[57] However, as the still significant welfare gains from the wage pooling policy (and the unchanged policy rankings) indicate, providing insurance against the quantitatively larger income risk is more beneficial, even when it comes with the reductions in incentives previously documented.[58] As before, due the complete deterioration of effort incentives, the combination of both policies is strongly dominated by the no wage discrimination policy.

7.1.3 Mismeasured Effort Inputs

The third key ingredient of our theory is the endogenous health effort choice that leads to better health outcomes in the future. We estimated the impact of effort on health transitions in section 5.2.2 under the assumption that our measures of health effort (the frequency of exercise and non-smoking behavior) are the only health-enhancing activities. Here we show that the presence of unobserved additional effort inputs can be modelled as non-classical measurement error in the effort variable e. Assume that true effort is given by

$$e = \lambda\tilde{e} + (1-\lambda)e^* \tag{24}$$

where $\lambda \in [0,1]$ is a parameter, $\tilde{e}$ is the effort we observe in the data and e^* are additional but unmeasured health effort inputs. Assume that the cost of producing true effort e is given by a CES function in $(\tilde{e}, e^*)$ with household-specific share and elasticity parameters $\chi = (\rho, \nu)$. Then the cost-minimizing choices of $(\tilde{e}, e^*)$ to achieve a given e are proportional to e (see Appendix I.4), and we can rewrite (24) as $e = \eta\lambda\tilde{e}$, where the random variable η has a cross-sectional distribution determined by the population distribution of the parameters (ρ, ν). If that distribution is such that η is a non-negative uniform random variable, symmetrically distributed around 1, with upper support[59] of $1/\lambda$, then $\eta \sim UNI[2-1/\lambda, 1/\lambda]$ and λ measures how noisy a proxy observed effort $\tilde{e}$ is for true effort e. If $\lambda = 1$, $\tilde{e}$ is a perfect proxy for e, as assumed thus far.

When we assume a measurement error of $\lambda = 0.75$ (the midpoint in the interval between no and maximal permissible error), in the re-estimated Q (displayed in Figure 21 in the appendix) health status updates respond moderately more to health effort today since households have more control over their health efforts than the noisy data suggest. This leads to larger adverse incentive effects from the policies, and thus somewhat lower welfare gains from them; see column QM in Table 10. The ranking of policies remains solidly intact, however, although the gap between the two policies shrinks, on account of the larger consequences from the under-provision of incentives in the wage non-discrimination policy.

[56] We now assume that insurance companies pool premia for each firm, but that firms do not discriminate with respect to hiring. Firms would not have an incentive to engage in cross-health type health expenditure subsidization that is socially efficient. Therefore, we assume that expenditures are chosen to maximize the expected net output of workers. This implies that the firms will choose the competitive expenditure schedule, $x^{CE}(\varepsilon, h)$ while the insurance premium P will be given by the analog of (4) under this expenditure rule. Because of this, labor earnings as a function of health will still be equal to the competitive level, $w^{CE}(h)$, but consumption is determined as $w^{CE}(h) - P$. Note that this equilibrium is distinct from the no prior equilibrium due to the difference in the $x(\varepsilon, h)$ schedule.

[57] The moderate gains from the no prior conditions policy stem from the fact that, under this policy, health expenditures are allocated efficiently, and in the premium pooling competitive equilibrium they are not.

[58] In the no wage discrimination policy, premium pooling is removed (but of course implemented if both policies are present).

[59] Since in our theory $e \in [0,1]$ and in the data $\tilde{e} \in [0,1]$ we require the upper point in the support of η to be $1/\lambda$, and by symmetry, the lower support of η to be $2 - 1/\lambda$, which is positive as long as $\lambda > 0.5$, which we will assume.

7.2 Sensitivity Analysis II: Measuring Health Expenditures and Their Impact

In our model the role of health expenditures is to determine health insurance premia. Thus far, we have assumed that empirically observed health expenditures represent real productivity-enhancing health expenditures that offset acute and productivity-reducing diseases and sicknesses. We now revisit this assumption.

7.2.1 Mismeasured Health Expenditures

First, one might be concerned that the observed cross-sectional dispersion in health expenditures partially stems from the dispersion of prices paid for the same services. We now argue that under at least one plausible set of assumptions about the health insurance market, this consideration leaves our analysis fundamentally unchanged. To this end, consider a simple variant of the model in which there is both a price (p) and a quantity (x) dimension to health care. Suppose insurance companies can specify the quantity of care $x(\varepsilon, h)$, but the price of this quantity is subject to a stochastic shock p, where $E(p) = 1$ and $p \sim \Pr(p)$. Further, assume that the insurance company only knows the distribution of p but not its realization when it contracts for real services $x(\varepsilon, h)$. Then p and x are *uncorrelated* and premia are given by:

$$P(h) = g(h)x(0,h) + (1 - g(h)) \int_0^{\bar{\varepsilon}} x(\varepsilon, h) f(\varepsilon) d\varepsilon \int_p p \Pr(p) dp.$$

Recalling that $E(p) = 1$ our theoretical analysis would go through completely unchanged. However, when taking the model to the data, the presence of price dispersion impacts the inferred distribution of real expenditures x and thus our estimated distribution of the ε-shock.

In our quantitative analysis, the ε-shock follows a log-normal distribution with age-varying mean and variance, $(\mu_\varepsilon(t), \sigma_\varepsilon(t))$. According to the model, real medical expenditures continue to be given by $x(t, educ, h, \varepsilon) = \max[0, \varepsilon - \bar{\varepsilon}(t, educ, h)]$. Thus, x follows a shifted log-normal distribution with mean $\mu_\varepsilon(t) - \bar{\varepsilon}(t, educ, h)$ and variance $\sigma_\varepsilon(t)$. Assuming that p is also log-normally distributed with mean $-\sigma_p^2/2$ and variance σ_p^2, the observed variance of nominal health expenditures now consists of both the variance in prices σ_p^2 and quantities x, σ_ε^2. So far, we have assumed that $\sigma_p^2 = 0$, but with price variance $\sigma_p^2 > 0$, part of the nominal expenditure variance in the data can be attributed to this channel. In fact, in our benchmark estimation the model struggles to explain the substantial expenditure variance in the data. Permitting positive price variance allows us to fit this moment without altering the other parameter estimates, and thus without changing any of the positive or normative findings from the benchmark model.

7.2.2 Misspecified Link Between Health Expenditures and Health

Even if we measure real health expenditures correctly in the data it might still be possible that we misspecify their impact on labor productivity. Thus, consider the other extreme: that health expenditures x are pure waste, or yield utility separable from consumption. Since the estimation of the competitive equilibrium would still match the same wage and health expenditure data, both the health-income gradient implied by the production function and the health-insurance premium gradient would remain intact. However, previously the equilibrium health expenditure profiles adjusted in response to the change in policy because of the productivity consequences of health expenditures. If these are absent, so will be the *endogenous response* of the health expenditure profile. In column FX of Table 10 we therefore repeat our analysis but keep the health insurance cutoffs unchanged at their competitive levels when introducing the policies. Both qualitatively and to a very large extent quantitatively, the policy conclusions remain unchanged under this alternative interpretation of the effects of health expenditures.[60] This finding suggests that the exact motivation for spending on health is not crucial for our policy results. Rather, what is crucial is that this motivation leads to the health expenditure differences by health status we observe in the data.

[60] Since the cutoffs are identical in the CE, the constrained efficient allocation and under both policies, there is no change in these rows. Only the cutoffs under the no prior conditions law are significantly different from the competitive benchmark (see Figure 11), and thus only under this policy does the fixing of the cutoffs have a non-trivial effect on the welfare results.

7.3 Sensitivity Analysis III: Modelling Uninsured Households

The purpose of this subsection is to demonstrate that the positive and normative implications of our model remain intact when it is extended to capture uninsured households. To do so, we now assume that an age-, health-status- and education-specific fraction of individuals cannot buy health insurance (but can purchase health goods out of pocket) for model-exogenous reasons.[61] Not being offered health insurance is *iid* over time, and the probability is determined by the empirical share of the uninsured.[62]

Since health spending affects labor productivity, we also have to specify the wage an uninsured household earns, as a function of her health spending x. We assume that the wage profile $w(h, \varepsilon)$ is the one offered by a production firm that takes the health expenditures of workers as given.[63] Consumption of an uninsured individual is then $c(h, \varepsilon) = w(h, \varepsilon) - x^{CE}(h, \varepsilon)$ and period utility of an uninsured is given by $U(h) = \int_{\varepsilon} u(c(h, \varepsilon))f(\varepsilon)d\varepsilon$. In the dynamic program, continuation utility now involves taking into account expectation over health insurance status, and the no prior conditions law is now interpreted to also provide insurance to all uninsured households. When repeating the policy analysis in this extended version of the model, we find, as summarized in column UN of Table 10, that the welfare benefits of the potential policy reforms become slightly larger due to the added insurance benefits for the uninsured.[64] The ε-induced risk is non-trivial and raises the attractiveness of the no prior conditions law, but not enough to change the policy rankings from the benchmark economy.

Why did we not make the model analyzed in this section the benchmark? In our model, in the absence of other inefficiencies in the health insurance market, all individuals should buy insurance (or have their employers provide it). Thus, we should only observe individuals with health insurance. To rationalize individuals without insurance, purchasing health insurance must be made inefficient for some of them. This can easily be achieved by introducing a (possibly age- health-status- and education-specific and possibly random) fixed cost. However, only a lower bound of these fixed costs (that makes individuals indifferent between buying and not buying insurance) can be identified from observed health insurance choices. Any higher fixed cost leads to the same choices. Thus, above the bound, the data are not informative about the fixed cost. However, our *normative* analysis of the no prior conditions law is sensitive to the magnitude of these fixed costs as they determine the net benefits of insuring individuals at this cost. Although the analysis conducted above reassures us that the broad conclusions we have obtained in the benchmark model without uninsured households is robust to their inclusion, we acknowledge-in fact we want to stress again explicitly-that our analysis should not be interpreted as a comprehensive investigation of all aspects of the ACA, and specifically not its attempt to reduce the share of the uninsured.

8 Conclusion

In this paper, we studied the effects of labor and health insurance market regulations on the evolution of health and production, as well as welfare. We showed that a no wage discrimination intervention in the labor market, combined with a no prior conditions intervention in the health insurance market provides effective consumption insurance against health shocks, holding the aggregate health distribution in society constant. However, the dynamic incentive costs and their impact on health and medical expenditures of both policies, if implemented jointly and perfectly, are large. As a result, providing partial but very significant wage insurance optimally balances this insurance-dynamic incentive trade-off. Furthermore, the optimal degree of social insurance in the labor market depends on the presence of social insurance in the health insurance market. More broadly, our paper therefore shows that a reliable policy analysis of health insurance reforms

[61] By simply forcing a fraction of individuals out of the market and ignoring the fixed costs necessary to rationalize this choice, we provide an upper bound for the benefits from a no prior conditions, mandatory health insurance law.

[62] We retain the assumption that the catastrophic health shocks z, remain fully covered even for those without health insurance, on account of the observation that hospitals have to treat even the uninsured for acute health conditions in their emergency rooms. The insured pay for this care through higher health insurance premia. The z component of insurance premia becomes $\mu_z(tc, educc, hc)/p(ins(tc, educc, hc))$, where $p(ins(tc, educc, hc))$ denote the measure of workers who are insured.

[63] In Proposition 16, equation (39) we construct a wage contract that induces individuals (in this case the uninsured) to spend $x = x^{CE}(h, \varepsilon)$ out-of-pocket exactly according to the competitive equilibrium schedule.

[64] Recall that the uninsured can still spend out of pocket, and the ε shocks are typically not large, the z shocks remain fully insured, and thus the impact of modelling the uninsured is quantitatively relatively small.

or labor market reforms cannot be conducted separately, since their interaction might deliver less favorable welfare results than those suggested by an isolated analysis of these policies.

We view several extensions as important for future work. First, the benefits of health in our model are confined to higher labor productivity, and have abstracted from a direct impact of better health on survival risk. Second, in our analysis labor income risk directly translates into consumption risk, in the absence of household private saving. We conjecture that the introduction of self-insurance via precautionary saving against this income risk further weakens the argument in favor of the policies studied in this paper. Third, in our model individuals can adjust health effort freely, although the high cost elasticity we estimate suggests that changing effort is not easy. Explicitly modeling addiction (for smoking) or habit persistence (for exercise or smoking) would likely delay the response of effort choices to policy incentives. Future work has to uncover whether such model extensions also impact our conclusions about the *relative* desirability of these policies.

References

[1] Acemoglu, D. and J. Angrist (2001), "Consequences of Employment Protection? The Case of the Americans with Disabilities Act." *Journal of Political Economy,* Volume 109, No. 5, 915-957.

[2] Aizawa, N. and H. Fang (2015), "Equilibrium Labor Market Search and Health Insurance Reform." *NBER Working Paper No. 18698.*

[3] Ales, L., R. Hosseini, and L. Jones (2014), "Is There "Too Much" Inequality in Health Spending Across Income Groups?" *NBER Working Paper 17937.*

[4] Attanasio, O., S. Kitao and G. Violante (2011), "Financing Medicare: A General Equilibrium Analysis." *Demography and the Economy* (chapter 9), J. Shoven (ed.), University of Chicago Press.

[5] Baicker, K., S. Taubman, H. Allen, M. Bernstein, J. Gruber, J. Newhouse, E. Schneider, B. Wright, A. Zaslavsky, and A.N. Finkelstein (2013), "The Oregon Health Insurance Experiment - Effects of Medicaid on Clinical Outcomes."*New England Journal of Medicine*, 368(18).

[6] Barbaresco, Silvia, Charles J. Courtemanche, Yangling Qi (2015), "Impacts of the Affordable Care Act dependent coverage provision on health-related outcomes of young adults," *Journal of Health Economics*, 40, 54-68.

[7] Bartel A. and P. Taubman (1979), "Health and Labor Market Success: The Role of Various Diseases."*Review of Economics and Statistics*, 61(1), 1-8.

[8] Befort, S. (2013), "An Empirical Examination of Case Outcomes under the ADA Amendments Act."*Washington and Lee Law Review*, 70(4), 2027-2071.

[9] Bhattacharya, J., K. Bundorf, N. Pace and N. Sood (2011), "Does Health Insurance Make You Fat?" In Grossman and Mocan (Eds.), *Economic Aspects of Obesity.* University of Chicago Press.

[10] Blundell, R., L. Pistaferri and I. Saporta-Eksten (2016), "Consumption Inequality and Family Labor Supply," *American Economic Review*, 106(2), 387-435.

[11] Booth, F., C. Roberts, and M. Laye (2012), "Lack of Exercise is a Major Cause of Chronic Diseases." *Comprehensive Physiology*, 2(2), 1143-1211.

[12] Braun, A., K. Kopecky and T. Koreshkova (2017), "Old, Sick, Alone and Poor: A Welfare Analysis of Old-Age Social Insurance Programs," *Review of Economic Studies*, 84, 580-612.

[13] Brügemann, B. and I. Manovskii (2010), "Fragility: A Quantitative Analysis of the US Health Insurance System," *Working paper*, University of Pennsylvania.

[14] Cawley, J. (2004), "The Impact of Obesity on Wages," *Journal of Human Resources*, 39(2), 451-474.

[15] Charles, Kerwin Kofi (2003), "The Longitudinal Structure of Earnings Losses among Work-Limited Disabled Workers," *Journal of Human Resources*, 38(3), 618-646.

[16] Charness, G. and U. Gneezy (2009), "Incentives to Exercise," *Econometrica*, 77, 909-931.

[17] Colman, G. and D. Dave (2012), "Physical Activity and Health," *International Journal of Arts & Sciences*, 5(6), 29-45.

[18] Currie, J. and B. Madrian (1999), "Health, Health Insurance and the Labor Market." In *Handbook of Labor Economics,* Volume 3C, edited by David Card and Orley Ashenfelter, Amsterdam: North Holland, 3309-3407.

[19] DeLeire, T. (2000) "The Wage and Employment Effects of the Americans with Disabilities Act." *Journal of Human Resources,* 35(4), 693-715.

[20] DeLeire, T. (2001) "Changes in Wage Discrimination against People with Disabilities: 1984-93." *Journal of Human Resources,* 35(1), 144-158.

[21] De Nardi, M., E. French and J. Jones (2016), "Medicaid Insurance in Old Age," *American Economic Review*, 106(11), 3480-3520.

[22] Dey, M. and C. Flinn (2005), "An Equilibrium Model of Health Insurance Provision and Wage Determination," *Econometrica* 73, 571-627.

[23] Ehrlich, I. and G. Becker (1972), "Market Insurance, Self-Insurance, and Self-Protection," *Journal of Political Economy*, 80, 623-648.

[24] Ehrlich, I. and H. Chuma (1990), "A Model of the Demand for Longevity and the Value of Life Extension," *Journal of Political Economy*, 98, 761-782.

[25] French, Eric (2005), "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behavior," *The Review of Economic Studies*, 72(2), 395-427.

[26] French, E. and J. Jones (2004), "On the Distribution and Dynamics of Health Care Costs," *Journal of Applied Economics*, 19, 705-721.

[27] French, E. and J. Jones (2011), "The Effects of Health Insurance and Self-Insurance on Retirement Behavior," *Econometrica*, 79(3), 693-732.

[28] Golosov, M. and A. Tsyvinski (2007), "Designing Optimal Disability Insurance: A Case for Asset Testing" *Journal of Political Economy*, 114, 257-279.

[29] Greenwood, J., Z. Hercowitz and G. Huffman (1988), "Investment, Capacity Utilization, and the Real Business Cycle" *American Economic Review*, 78, 402-17.

[30] Grossman, M. (1972), "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80(2), 223-255.

[31] Hall, R. and C. Jones (2007), "The Value of Life and the Rise in Health Spending," *Quarterly Journal of Economics*, 122(1), 39-72.

[32] Halliday, T., H. He, L. Ning, and H. Zhang (2016), "Health Investment over the Life Cycle," *Macroeconomic Dynamics.*

[33] Hamman, RF, RR Wing, SL Edelstein, JM Lachin, GA Bray, L. Delahanty, M. Hoskin, AM Kriska, EJ Mayer-Davis, X. Pi-Sunyer, J. Regensteiner, B. Venditti, and J. Wylie-Rosett (2006), "Effect of weight loss with lifestyle intervention on risk of diabetes," *Diabetes Care*, 29(9), 2102-7.

[34] Hansen, G., M. Hsu and J. Lee (2014), "Health Insurance Reform: The Impact of a Medicare Buy-In," *Journal of Economic Dynamics and Control*, 45(C), 315-329.

[35] Hopenhayn, H. and J. Nicolini (1997), "Optimal Unemployment Insurance." *Journal of Political Economy*, 105, 412-438.

[36] Hugonnier, J., F. Pelgrin and P. St-Amour (2013), "Health and (Other) Asset Holdings," *Review of Economic Studies*, 80(2), 663-710.

[37] Jeske, K. and S. Kitao (2009), "U.S. Tax Policy and Health Insurance Demand: Can a Regressive Policy Improve Welfare?," *Journal of Monetary Economics*, 56, 210-221.

[38] Jung, J. and C. Tran (2016), "Market Inefficiency, Insurance Mandate and Welfare: U.S. Health Care Reform 2010," *Review of Economic Dynamics*, 20, 132-159.

[39] Kowalski, A. (2015), "Estimating the Tradeoff Between Risk Protection and Moral Hazard with a Nonlinear Budget Set Model of Health Insurance," *International Journal of Industrial Organization*, 43, 122-135.

[40] Low, H. and L. Pistaferri (2015), "Disability Insurance and the Dynamics of the Incentive Insurance Trade-Off," *American Economic Review*, 105, 2986-3029.

[41] Mirrlees, J. (1971) "An Exploration in the Theory of Optimum Income Taxation," *Reivew of Economic Studies*, 38, 175-208.

[42] Mitchell, J. and R.V. Burkhauser (1990). "Disentangling the Effect of Arthritis on Earnings: A Simultaneous Estimate of Wage Rates and Hours Worked," *Applied Economics*, 22(10), 1291-1309.

[43] Mitchell, J. and J.S. Butler (1986), "Arthritis and the Earnings of Men: An Analysis Incorporating Selection Bias," *Journal of Health Economics*, 5(1), 81-98.

[44] O'Donnell, O., E. Van Doorslaer, and T. Van Ourti (2015), "Health and Inequality." In A.B. Atkinson and F. Bourguignon (Eds.), Handbook of Income Distribution volume 2B (1419-1533). Amsterdam: Elsevier B.V.

[45] OECD (2010), "Sickness, Disability and Work: Breaking the Barriers."

[46] Ozkan, S. (2014), "Preventive vs. Curative Medicine: A Macroeconomic Analysis of Health Care over the Life Cycle," *Working Paper*.

[47] Pashchenko, S. and P. Porapakkarm (2013), "Quantitative Analysis of Health Insurance Reform: Separating Regulation from Redistribution," *Review of Economic Dynamics*, 16, 383-404.

[48] Pelkowski, Jodi Messer and Mark C. Berger (2004), "The Impact of Health on Employment, Wages, and Hours Worked over the Life Cycle," *The Quarterly Review of Economics and Finance*, 44: 102-121.

[49] Pijoan-Mas, J. and V. Rios-Rull (2014), "Heterogeneity in Expected Longevities," *Demography*, 51(6), 2075-2102.

[50] Shavell, S. and L. Weiss (1979), "The Optimal Payment of Unemployment Insurance Benefits over Time." *Journal of Political Economy*, 87, 1347-1362.

[51] Simon, K., A. Soni, and J. Cawley (2017), "The Impact of Health Insurance on Preventive Care and Health Behaviors: Evidence from the First Two Years of the ACA Medicaid Expansions," *Journal of Policy Analysis and Management*, 36(2), 390-417.

[52] Tuomilehto J., J. Lindstrm, JG Eriksson, TT Valle, H. Hmlinen, P. Ilanne-Parikka, S. Keinnen-Kiukaanniemi, M. Laakso, A. Louheranta, M. Rastas, V. Salminen, M. Uusitupa; Finnish Diabetes Prevention Study Group (2001), "Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance," *The New England Journal of Medicine*, 344(18), 1343-50.

Appendix for Referees and Online Publication

A Social Insurance Policies for Health across Space

In 2010, the OECD published a comprehensive report entitled "Sickness, Disability and Work: Breaking the Barriers," which summarizes policies for disabled individuals and analyzes the recent policy reforms in the OECD countries. As a summary measure of understanding different policies in place, the report generates two policy indicators: the compensation index and the integration (and employment) index. While the former captures the generosity of policies like disability benefits, the latter is focused on understanding the degree to which the government helps the disabled in the labor market. Some of the policies included in the integration index are: anti-discrimination legislation; employment quotas; employer incentives; supported employment; and wage subsidies. According to the report, every single country improved its integration measures from 1990 to 2007, whereas there was, on average, a drop in the compensation index over the same period. On average, the integration index increased 10 points on a 50-point scale, with the Netherlands, the United Kingdom, Finland, and Australia recording an increase of over 15 points. Thus, the OECD countries have strengthened, in one way or another, labor market policies for disabled workers in the recent decade. They have, in many cases, broadened the definition of disability, implying that there are more labor market protections for unhealthy workers in general.

In Table 11, we summarize the social insurance policies in the health insurance market and the labor market for several countries, including the G7 (Canada, France, Germany, Italy, Japan, the United Kingdom and the United States).

Table 11: Social Insurance Policies against Health Risks across Space

Country	Health Insurance	Labor Market
Canada	• Medicare (publicly funded)	• Canadian Charter Rights and Freedoms (1985); Employment Equity Act (1995)
	• Health-risk-related factors are not permitted in premium setting in private markets	• Prohibits discrimination and needs to make reasonable accommodations
France	• National health insurance	• Disability Employment Act (1987), amended in 2005 • Employment quota system
Germany	• Compulsory enrollment into public "sickness funds" at common rates for all	• Employment quota system (employers with more than 20 employees assign 5% of the jobs to the disabled, with fines if violated) and special dismissal protection
	• Complementary private insurance	• Wage subsidies, financial support for workplace adaptation, tax reduction and special arrangements for sick leave
Italy	• National health service	• Normative for Employment Law for the Disabled (1999) • Employment quota system

Japan	• Employees' Health Insurance (for workers) and National Health Insurance (for students and self-employed)	• Act on Employment Promotion of Persons with Disabilities (1960, with most recent amendments in 2005) • Employment quota system: expanded to include smaller firms and broadened the definition of disability
United Kingdom	• National Health Service - publicly funded health care	• Equality Act 2010 (replaced Disability Discrimination Act of 1995) • Employers required to make reasonable adjustments to workplaces
United States	• Patient Protection and Affordable Care Act (2010)	• Americans with Disabilities Act (1990) and its Amendments (2008)
Sweden	• Government-funded health care	• Discrimination Act (2009) strengthened protection • Wages are regulated by collective agreements (and thus disabled workers are remunerated according to the agreement) • Flexible working hours, technical aids and adaptations provided to accommodate the needs of disabled workers • Employer receives wage subsidies
Switzerland	• Compulsory basic insurance in private market, whose premia must be equal for everyone	• Federal Act on the Elimination of Discrimination against People with Disabilities (EPDA)

Note: In general, the Employment Equality Framework Directive is one of the employment discrimination laws in the European Union that prohibit discrimination on the basis of age, disability, sexual orientation, and religion or belief in the workplace. *Source*: Colombo and Tapay (2004); European Commission (2012); Hasegawa (2010); OECD (2010).

B Proofs of Propositions

Proposition 5

Proof. Since exercise does not carry any benefits in the static model, trivially $e^{SP} = 0$. Attaching he Lagrange multiplier $\mu \geq 0$ to the resource constraint, the first-order condition with respect to consumption $c(\varepsilon)$ is

$$u'(c(\varepsilon, h)) = \lambda$$

and thus $c^{SP}(\varepsilon, h) = c^{SP}$ for all $\varepsilon \in E$ and $h \in H$. Thus, not surprisingly, the social planner provides full consumption insurance to households. The optimal health expenditure allocation maximizes this consumption

$$c^{SP} = \max_{x(\varepsilon,h)} \sum_h \left\{ g(h) \left[F(h, -x(0,h)) - x(0,h) \right] + (1 - g(h)) \int f(\varepsilon) \left[F(h, \varepsilon - x(\varepsilon, h)) - x(\varepsilon, h) \right] d\varepsilon \right\} \Phi(h)$$

Denoting by $\mu(\varepsilon, h) \geq 0$ the Lagrange multiplier on the constraint $x(\varepsilon, h) \geq 0$, the first-order condition with respect to $x(\varepsilon, h)$ reads as

$$-F_2(h, \varepsilon - x(\varepsilon, h)) + \mu(\varepsilon, h) = 1$$

Fix $h \in H$. By assumption 4 $F_{22}(h, y) < 0$ and thus either $x(\varepsilon, h) = 0$ or $x(\varepsilon, h) > 0$ satisfying

$$-F_2(h, \varepsilon - x(\varepsilon, h)) = 1$$

for all ε. Thus off-corners $\varepsilon - x(\varepsilon, h) = \bar{\varepsilon}^{SP}(h)$ where the threshold satisfies

$$-F_2(h, \bar{\varepsilon}^{SP}(h)) = 1. \tag{25}$$

Consequently

$$x^{SP}(\varepsilon, h) = \max\left[0, \varepsilon - \bar{\varepsilon}^{SP}(h)\right].$$

The fact that $\bar{\varepsilon}^{SP}(h)$ is increasing in h, strictly so if $F_{12}(h, y) > 0$, follows directly from assumption 4 and (25). ■

Proposition 6

Proof. Attaching he Lagrange multiplier $\mu(h)$ to equation (2) and $\lambda(h)$ to equation (3) the first-order conditions read as

$$\begin{aligned}
u'(w(h) - P(h)) &= \lambda(h) = -\mu(h) && (26)\\
\lambda(h)F_2(h, -x(0, h)) &\leq \mu(h) && (27)\\
&= \text{ if } x(0, h) > 0 &&\\
\lambda(h)F_2(h, \varepsilon - x(\varepsilon, h)) &\leq \mu(h) && (28)\\
&= \text{ if } x(\varepsilon, h) > 0 &&
\end{aligned}$$

Thus off-corners we have

$$F_2(h, \hat{\varepsilon} - x(\hat{\varepsilon}, h)) = F_2(h, \varepsilon - x(\varepsilon, h)) = K \tag{29}$$

for some constant K. Thus off-corners $\varepsilon - x(\varepsilon, h)$ is constant in ε and thus medical expenditures satisfy the cutoff rule

$$x^{CE}(\varepsilon, h) = \max\left[0, \varepsilon - \bar{\varepsilon}^{CE}(h)\right]. \tag{30}$$

Plugging (30) into (28) and evaluating it at $\varepsilon = \bar{\varepsilon}^{CE}(h)$ yields

$$\lambda(h)F_2(h, \bar{\varepsilon}^{CE}(h)) = \mu(h). \tag{31}$$

Using this result in the second part of (26) delivers the characterization of the equilibrium cutoff levels

$$F_2(h, \bar{\varepsilon}^{CE}(h)) = -1 \text{ for all } h \in H$$

which are unique, given the assumptions imposed on F. Wages, consumption and health insurance premia then trivially follow from (2) and (3). ■

Proposition 16

Proof.

Suppose that the worker was offered compensation $\boldsymbol{w}(h, \varepsilon - x)$ as a function of his health status h and his productivity as given in equation (39). Then, note from that the worker's insurance choice problem can be written as

$$\begin{aligned}
&\max_{x(\varepsilon,h)\geq 0} \int u\left[\boldsymbol{w}(h, \varepsilon - x(\varepsilon, h)) - P(h)\right] f(\varepsilon) d\varepsilon \\
&s.t. \\
P(h) &= \int x(\varepsilon, h) f(\varepsilon) d\varepsilon
\end{aligned}$$

From (39) we observe that wages are constant for all $x \geq \varepsilon - \bar{\varepsilon}^{CE}(h)$ and thus $x(\varepsilon, h) \leq \varepsilon - \bar{\varepsilon}^{CE}(h)$ for all ε. Since $x(\varepsilon, h)$ is restricted to be non-negative it follows that $x(\varepsilon, h) = 0$ for all $\varepsilon \leq \bar{\varepsilon}^{CE}(h)$. Conditional on $x(\tilde{\varepsilon}, h) > 0$ for a given shock $\tilde{\varepsilon} > \bar{\varepsilon}^{CE}(h)$ the first-order conditions read as

$$\begin{aligned} u'\left[\boldsymbol{w}(h, \tilde{\varepsilon} - x(\tilde{\varepsilon})) - P(h)\right] f(\tilde{\varepsilon}) \frac{\partial \boldsymbol{w}(h, \tilde{\varepsilon} - x(\tilde{\varepsilon}))}{\partial x(\tilde{\varepsilon}, h)} &= f(\tilde{\varepsilon})\lambda(h) \\ \int u'\left[\boldsymbol{w}(h, \varepsilon - x(\varepsilon, h)) - P(h)\right] f(\varepsilon) d\varepsilon &= \lambda(h) \end{aligned}$$

Combining, simplifying and exploiting the fact that $\frac{\partial \boldsymbol{w}(h, \tilde{\varepsilon} - x(\tilde{\varepsilon}))}{\partial x(\tilde{\varepsilon}, h)} = -F_2(h, \varepsilon - x(\varepsilon, h))$ for $x(\varepsilon, h) \leq \varepsilon - \bar{\varepsilon}^{CE}(h)$ yields

$$-F_2(h, \varepsilon - x(\varepsilon, h)) = \frac{\int u'\left[\boldsymbol{w}(h, \varepsilon - x(\varepsilon, h)) - P(h)\right] f(\varepsilon) d\varepsilon}{u'\left[\boldsymbol{w}(h, \tilde{\varepsilon} - x(\tilde{\varepsilon})) - P(h)\right]}. \tag{32}$$

But the health expenditure allocation

$$x(\varepsilon, h) = \max[0, \varepsilon - \bar{\varepsilon}^{CE}(h)]$$

yields

$$\boldsymbol{w}(h, \varepsilon - x(\varepsilon, h)) = w^{CE}(h)$$

for all ε, and thus (32) becomes

$$-F_2(h, \bar{\varepsilon}^{CE}(h)) = 1$$

which is satisfied by the definition of $\bar{\varepsilon}^{CE}(h)$. Next, note that the firm's profits under the pooled wage-health insurance contract is

$$F(h, \varepsilon - x) - w^{CE}(h) = \begin{cases} F(h, \varepsilon) - w^{CE}(h) & \text{if } \varepsilon < \bar{\varepsilon}^{CE}(h) \\ F(h, \bar{\varepsilon}^{CE}(h)) - w^{CE}(h) & \text{otherwise} \end{cases}. \tag{33}$$

Now we show that under the proposed separation of wage and health insurance contracts the production firm does not need to know whether the worker has purchased health insurance. Small shocks $\varepsilon < \bar{\varepsilon}^{CE}(h)$ are not covered in any case, and for shocks $\varepsilon \geq \bar{\varepsilon}^{CE}(h)$ net profits of the firm with worker insurance are given as

$$F(h, \bar{\varepsilon}^{CE}(h)) - w^{CE}(h)$$

and without insurance

$$\begin{aligned} & F(h, \varepsilon - x) - \boldsymbol{w}(h, \varepsilon - x) \\ = \; & F(h, \varepsilon - x) - \left\{ w^{CE}(h) - \left[F(h, \bar{\varepsilon}^{CE}(h) - F(h, \varepsilon - x) \right] \right\} \\ = \; & F(h, \bar{\varepsilon}^{CE}(h)) - w^{CE}(h) \end{aligned}$$

Thus net profits of the firm are the same whether or not the worker buys insurance for health shocks $\varepsilon \geq \bar{\varepsilon}^{CE}(h)$, vacating the need for the firm to verify whether the worker has purchased health insurance elsewhere. ■

Proposition 7

Proof. Let the Lagrange multipliers to equations (4) and (3) be μ and $\lambda(h)$, respectively. Then, the first-order conditions are:

$$\begin{aligned} \sum_h u'(w(h) - P)\Phi(h) &= \mu \\ u'(w(h) - P)\Phi(h) &= \lambda(h) \\ (1 - g(h))f(\varepsilon)[-F_2(h, \varepsilon - x(\varepsilon, h))]\lambda(h) &\leq \mu(1 - g(h))f(\varepsilon)\Phi(h) \\ &= \text{ if } x(\varepsilon, h) > 0 \\ g(h)[-F_2(h, -x(0, h))]\lambda(h) &\leq \mu g(h)\Phi(h) \\ &= \text{ if } x(0, h) > 0 \end{aligned}$$

Thus, off-corners we have

$$F_2(h, \varepsilon - x(\varepsilon, h)) = F_2(h, \hat{\varepsilon} - x(\hat{\varepsilon}, h)) = K$$

for some constant K and the cutoff rule is determined by

$$u'(w(h) - P)[-F_2(h, \bar{\varepsilon}^{NP}(h))] = \sum_h u'(w(h) - P)\Phi(h). \tag{34}$$

Moreover, let us take the derivative of (34) with respect to h.

$$\begin{aligned}
& u''(w(h) - P)\frac{\partial w(h)}{\partial h}F_2 + u'(w(h) - P)\left\{F_{12} + F_{22}\frac{\partial \bar{\varepsilon}^{NP}(h)}{\partial h}\right\} = 0 \\
& u''(w(h) - P)\frac{\partial \bar{\varepsilon}^{NP}(h)}{\partial h}\frac{\partial w(h)}{\partial \bar{\varepsilon}^{NP}(h)}F_2 + u'(w(h) - P)\left\{F_{12} + F_{22}\frac{\partial \bar{\varepsilon}^{NP}(h)}{\partial h}\right\} = 0 \\
\Rightarrow \quad & \frac{\partial \bar{\varepsilon}^{NP}(h)}{\partial h}\left\{u''(w(h) - P)F_2\frac{\partial w(h)}{\partial \bar{\varepsilon}^{NP}(h)} + u'(w(h) - P)F_{22}\right\} = -u'(w(h) - P)F_{12}
\end{aligned}$$

Note that as $\bar{\varepsilon}$ increases $w(h)$ decreases, since $F(h, \varepsilon - x(\varepsilon, h))$ is decreasing for $\varepsilon < \bar{\varepsilon}$, and constant for $\varepsilon \geq \bar{\varepsilon}$. Thus, we have

$$\frac{\partial \bar{\varepsilon}^{NP}(h)}{\partial h} > 0.$$

■

Proposition 8

Proof. From (9), we immediately obtain

$$-F_2(h, \bar{\varepsilon}^{NP}(h)) = \frac{\sum u'(w(h) - P)\Phi(h)}{u'(w(h) - P)} \begin{array}{c} < 1 \\ = 1 \\ > 1 \end{array} \Rightarrow \begin{array}{c} \bar{\varepsilon}^{NP}(h) < \bar{\varepsilon}^{SP}(h) \\ \bar{\varepsilon}^{NP}(h) = \bar{\varepsilon}^{SP}(h) \\ \bar{\varepsilon}^{NP}(h) > \bar{\varepsilon}^{SP}(h) \end{array}$$

as $-F_2(h, \bar{\varepsilon}^{SP}(h)) = 1$.

Let us take $h_L < \tilde{h} < h_H$, and suppose

$$-F_2(h_L, \bar{\varepsilon}^{NP}(h_L)) > 1 > -F_2(h_H, \bar{\varepsilon}^{NP}(h_H)), \tag{35}$$

i.e.

$$\begin{array}{rcl} \bar{\varepsilon}^{NP}(h_H) < \bar{\varepsilon}^{SP}(h_H) & \Rightarrow & w^{NP}(h_H) > w^{SP}(h_H) \\ \bar{\varepsilon}^{NP}(h_L) > \bar{\varepsilon}^{SP}(h_L) & \Rightarrow & w^{NP}(h_L) < w^{SP}(h_L), \end{array}$$

where $w^{SP}(h) = g(h)F(h, 0) + (1 - g(h)) \int f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon$. Then, we have

$$u'^{NP}(c(h_H) - P) < u'^{SP}(c(h_H) - P) < u'^{SP}(c(h_L) - P) < u'^{NP}(c(h_L) - P),$$

where the second inequality follows from (38). This result, in combination with (35), implies

$$u'^{NP}(c(h_L) - P)[-F_2(h_L, \bar{\varepsilon}^{NP}(h_L))] > u'^{NP}(c(h_H) - P)[-F_2(h_H, \bar{\varepsilon}^{NP}(h_H))],$$

a contradiction to (9). ■

Proposition 11

Proof. The fact that $e_T(h) = 0$ and $c_0(h) = c_0$ follows trivially from the fact that terminal effort has no positive consequence, and initial consumption dispersion has no positive incentive effect on effort. For all other effort and consumption levels, note that in the *iid* case, the social planner's problem becomes effectively static with current effort only affecting tomorrow's health distribution and, through that, tomorrow's consumption. With this we can write the problem as

$$\max_{c(h),e} \beta \sum_{h} u\left(c(h')\right) Q(h';e) - q(e) \qquad s.t.$$

$$\begin{aligned}
\sum_{h'} c(h')Q(h';e) &\leq \sum_{h'} \left[w^{SP}(h') - P^{SP}(h')\right] Q(h';e) \\
q'(e) &= \beta \sum_{h'} \frac{\partial Q(h';e)}{\partial e} u\left(c(h')\right).
\end{aligned}$$

We characterized $(w^{SP}(h'), P^{SP}(h'))$ in Proposition 5 and noted in the discussion of the static competitive equilibrium that $w^{SP}(h') - P^{SP}(h')$ is increasing in h' (since the planner is maximizing this object and since fixing the expenditure policy, production is increasing and expenditures are decreasing in health).

The first-order condition for consumption is

$$\beta u'\left(c(h')\right) \left[Q(h';e) - \lambda \frac{\partial Q(h';e)}{\partial e}\right] - \phi Q(h';e) = 0.$$

This implies that there is no consumption dispersion if $\lambda = 0$, and that consumption is ranked by h if $\lambda > 0$ since we have assumed that $\partial Q/\partial e$ is increasing in h'.

When $\lambda = 0$ and there is no consumption dispersion, then $e = 0$ from the effort IC condition. Moreover, if $\lambda < 0$ then consumption is declining in health and this too will imply that $e = 0$, but then the planner can do strictly better by giving everyone the average level of consumption. Hence, the effort condition can only bind from below, meaning that $\lambda \geq 0$. The first-order condition for effort is

$$\begin{aligned}
0 &= \beta \sum_{h'} \frac{\partial Q(h';e)}{\partial e} u\left(c(h')\right) - q'(e) \\
&+ \phi \sum_{h'} \left[w^{SP}(h') - P^{SP}(h') - c(h')\right] \frac{\partial Q(h';e)}{\partial e} \\
&+ \lambda \left[q''(e) - \beta \sum_{h'} \frac{\partial^2 Q(h';e)}{\partial e^2} u\left(c(h')\right)\right]
\end{aligned}$$

This simplifies to the following since the f.o.c. for effort implies that the direct effect on the objective is 0,

$$\begin{aligned}
0 &= \phi \sum_{h'} \left[w^{SP}(h') - P^{SP}(h') - c(h')\right] \frac{\partial Q(h';e)}{\partial e} \\
&+ \lambda \left[q''(e) - \beta \sum_{h'} \frac{\partial^2 Q(h';e)}{\partial e^2} u\left(c(h')\right)\right].
\end{aligned}$$

The resource constraint must bind, so $\phi > 0$. If $\lambda = 0$, we get a contradiction since the first term is positive. Hence, $\lambda > 0$ because

$$q''(e) - \beta \sum_{h'} \frac{\partial^2 Q(h';e)}{\partial e^2} u\left(c(h')\right) > 0$$

and the incentive constraint binds. This implies from the effort condition that $c(h')$ is positively ranked by h', and this in turn implies that $e > 0$ from the effort IC condition since $q'(e) = 0$. ■

Proposition 12

Proof. Is by backward induction. Trivially $e_T(h) = 0$. In period T, since both policies are in place, the wage and health insurance premium of every household are independent of h. Thus

$$v_T(h) = u(w_T - P_T) = v_T$$

and therefore the terminal value function is independent of h. Now suppose for a given time period t the value function v_{t+1} is independent of h. Then from the first-order condition with respect to $e_t(h)$ we have

$$q'(e_t(h)) = \beta v_{t+1} \sum_{h'} \frac{\partial Q(h'; h, e)}{\partial e}$$

But since for every e and every h, $Q(h'; h, e)$ is a probability measure over h' we have $\sum_{h'} \frac{\partial Q(h'; h, e)}{\partial e} = 0$ and thus $e_t(h, \gamma) = 0$ for all h, on account of our assumptions on $q'(.)$. But then

$$v_t(h) = u(w_t - P_t) + \left\{ -0 + \beta v_{t+1} \sum_{h'} Q(h'; h, 0) \right\} = u(w_t - P_t) + \beta v_{t+1} = v_t$$

since $\sum_{h'} Q(h'; h, 0) = 1$ for all h. Thus v_t is independent of h. The evolution of the health distributions follows from (11), and given these health distributions wages and health insurance premia are given by (18) and (20). ■

C Further Analysis of the No Wage Discrimination Case

C.1 Health Insurance Distortions with No Wage Discrimination

The firm's break-even condition is

$$\sum_h \left\{ g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)[F(h, \varepsilon - x^{NP}(\varepsilon, h))]d\varepsilon - w(h) \right\} \Phi(h) = 0,$$

and hence on average the production level of a worker will equal his gross wage. Taking $\varepsilon_w > 0$ and $\delta > 0$ as given, workers for whom the wage limits, $\max_{h,h'} |w(h) - w(h')| \leq \varepsilon_w$, bind will be paid either more or less than their production level depending on whether the wage discrimination bound binds from above or below. The firm will optimally choose to hire less than the population share of any health type h whose wage is above their production level, and hence some of these workers will be unemployed. Since we have assumed that there is no cost to working and workers pay for their own insurance, competition over health insurance will lead these workers to increase their health insurance, $x(e, h)$, so that their productivity is within ε_w of their wage $w(h)$. In the limit as $\varepsilon_w \to 0$, this implies that

$$w(h) = g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)[F(h, \varepsilon - x^{NP}(\varepsilon, h))]d\varepsilon, \tag{36}$$

holds and they are fully employed, or $w(h) - P(h) = 0$. On the flip side, since there will be excess demand for workers whose expected production is more than $w(h)$, they will therefore find it optimal to lower their insurance, and in the limit as $\varepsilon \to 0$, either (36) holds or they set $x(e, h) = 0$ if the workers end up at a corner with respect to health insurance. Assuming that neither corner binds, this implies that the no wage discrimination policy will be undone by adjustments in the health insurance market. This motivated our assumption that the government will choose to regulate the health insurance market to prevent this outcome as part of the no wage discrimination policy.

For health types for which the bounds do not bind, market clearing implies that

$$w(h) = g(h)F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)[F(h, \varepsilon - x^{NP}(\varepsilon, h))]d\varepsilon$$

while actuarial fairness implies that

$$P(h) = (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon)x^{NP}(\varepsilon, h))d\varepsilon.$$

Hence, an efficient health insurance contract for this type will maximize $w(h) - P(h) = w^{CE}(h) - P^{CE}(h)$. Since $w^{CE}(h) - P^{CE}(h)$ is increasing in h, it follows that the wage bound binds for the lowest and highest health types.

C.2 No Wage Discrimination with Realized Penalties in Equilibrium

Here we assume that the firm must pay a cost for having wage dispersion conditional on health type or for having the health composition of its work force differ from the population average. The wage variation penalty is assumed to take the form

$$C\sum_h [w(h) - w(0)]^2\, n(h),$$

since health type 0 will have the lowest wage in equilibrium, and where C is the penalty parameter and $n(h)$ is a measure of type h workers the firm hires. Note that with this penalty function, the penalty will apply to all workers with health $h > 0$. The penalty from having one's composition deviate from the population average is given by

$$\sum_h D\left[\frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum \Phi(h)}\right]^2.$$

Since these penalties are small for small deviations, it will turn out that penalty costs will be realized in equilibrium. Since both of these penalties are real we need to subtract them from production. We will assume that here too the government will regulate the insurance market to prevent low health status workers from raising their productivity by over-insuring themselves against health risks and high health status workers from lowering their productivity by under-insuring themselves.

We begin analyzing this case by assuming that the penalties for wage discrimination C and hiring discrimination D are both finite and then we examine the equilibrium in the limit as they become large. The firm takes as given the health policy of the worker and the equilibrium wage $w(h)$ and chooses the measure of each health type to hire $n(h)$ so as to maximize

$$\begin{aligned}&\max_{n(h)}\sum_h\left[g(h)\left[F(h,-x(0,h)) - x(0,h)\right] + (1-g(h))\int_0^{\bar{\varepsilon}} f(\varepsilon)\left[F(h,\varepsilon - x(\varepsilon,h)) - x(\varepsilon,h)\right]d\varepsilon - w(h)\right]n(h)\\ &-C\sum_h [w(h)-w^*]^2\, n(h) - \sum_h\left[\frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum\Phi(h)}\right]^2,\end{aligned}$$

where w^* is taken here to mean the lowest wage. Trivially, the firm will want to hire more than the population share of any type h for whom

$$\begin{aligned}N(h) \quad\equiv\quad &\left[g(h)\left[F(h,-x(0,h)) - x(0,h)\right] + (1-g(h))\int_0^{\bar{\varepsilon}} f(\varepsilon)\left[F(h,\varepsilon - x(\varepsilon,h)) - x(\varepsilon,h)\right]d\varepsilon - w(h)\right]\\ &-C\left[w(h)-w^*\right]^2\end{aligned}$$

is positive and less than the population share if $N(h)$ is negative. Since all firms share this condition, they will all choose the same relative shares of each type of worker. Since workers are willing to work so long as $w(h) - P(h) > 0$, it follows that $w(h)$ cannot be more than w^* if $N(h)$ is not positive. To see this, note that there would be an excess supply of type h workers and hence the labor market would not clear. Moreover, a firm would rather hire a worker of type h at $w^* - \varepsilon$ than for w^* for ε small. Hence, if $w(h) = w^*$, then $N(h) = 0$ so long as $w^* - P(h) > 0$. Hence, for the labor market to clear for each health type, either $N(h) = 0$ for type h or $N(h) > 0$ but $w(h) - P(h) = 0$. Since the government can set $x(\varepsilon, h) = 0$, which implies that $P(h) = 0$, we assume that $w(h) - P(h) > 0$ for all health types. This implies the following proposition.

Proposition 13 *If C and D are positive but finite, and $w(h) - P(h) > 0$ for all h, then in equilibrium all households are hired, all firms are representative, and the wage $w(h)$ is equal to a worker's productivity less the cost of paying him. As C gets large, $w(h)$ converges to w^* for all $h > 0$, and the health-related productivity differences are consumed by the enforcement costs.*

C.3 Realized Penalties with Both Policies

Since all that workers care about is their net wage $\tilde{w}(h)$, which is also equal to their consumption, it follows that workers are indifferent over contracts that offer combinations of a gross wage $w(h)$ and medical costs

$P(h)$ for which $\tilde{w}(h) = w(h) - P(h)$ is constant. Hence, it is natural to assume that the firm takes the equilibrium *net wage* function $\tilde{w}(h)$ as given and chooses the measure of each health type to hire, $n(h)$, and its health plan, $x(\varepsilon, h)$, to solve the following problem

$$\max_{n(h),x(\varepsilon,h)} \sum_h \left[g(h) \left[F(h, -x(0,h)) - x(0,h) \right] + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) \left[F(h, \varepsilon - x(\varepsilon, h)) - x(\varepsilon, h) \right] d\varepsilon - \tilde{w}(h) \right] n(h)$$

$$-C \sum_h \left[\tilde{w}(h) - \tilde{w}(0) \right]^2 n(h) - \sum_h D \left[\frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum \Phi(h)} \right]^2.$$

Proposition 14 *If C and D are positive but finite, then in equilibrium all households are hired, all firms are representative, the net wage $\tilde{w}(h)$ is equal to a worker's productivity less the cost of paying him more than $\tilde{w}(0)$, and $\tilde{w}(0) = w^{CE}(0) - P(0)$. The firm optimally sets $x(\varepsilon, h) = x^{CE}(\varepsilon, h)$. As $C \to \infty$, $\tilde{w}(h) \to \tilde{w}(0)$.*

Proof. The optimality condition for $x(h, \varepsilon)$ if $\varepsilon = 0$ is

$$F(h, -x(0,h)) - 1 \leq 0$$

and if $\varepsilon > 0$ is

$$F(h, \varepsilon - x(\varepsilon, h)) - 1 \leq 0 \text{ w. equality if } x(\varepsilon, h) > 0.$$

These are the same conditions as in the competitive equilibrium.

Next, we show that $\tilde{w}(h)$ has to be increasing in h and hence $\tilde{w}(0)$ is the lowest paid type. The wage penalty is w.r.t. to the lowest paid worker type, which we denote by w^*. Given that optimum insurance is the same as in the competitive equilibrium, it follows that net earnings per worker is $w^{CE}(h) - P^{CE}(h) - \tilde{w}(h)$, and from before $w^{CE}(h) - P^{CE}(h)$ is increasing in h. Hence, for the firm to break even

$$\begin{aligned} & \sum_h \left[w^{CE}(h) - P^{CE}(h) - \tilde{w}(h) \right] n(h) \\ - \quad & C \sum_h \left[\tilde{w}(h) - w^* \right]^2 n(h) - \sum_h D \left[\frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum \Phi(h)} \right]^2 = 0, \end{aligned}$$

and the optimality condition for $n(h)$ is

$$\begin{aligned} & \left[w^{CE}(h) - P^{CE}(h) - \tilde{w}(h) \right] - C \left[\tilde{w}(h) - w^* \right]^2 \\ - \quad & D \left[\frac{n(h)}{\sum n(h)} - \frac{\Phi(h)}{\sum \Phi(h)} \right] \left[1 - \frac{n(h)}{\sum n(h)} \right] \frac{1}{\sum n(h)} = 0. \end{aligned}$$

This condition implies that a firm will hire more than the population share of any type h for whom

$$\tilde{N}(h) \equiv w^{CE}(h) - P^{CE}(h) - \tilde{w}(h) - C \left[\tilde{w}(h) - w^* \right]^2 > 0,$$

and less than the population share if the reverse is true. However any health type h who are not fully employed in equilibrium would have excess members who would be happy to be hired at any positive wage. Hence, either type h is paid the lowest equilibrium wage or they are fully employed. Hence, any type h for whom $w(h) > w^*$ are fully employed. Any type receiving the lowest wage must be fully employed since the firm would be willing to hire more of these workers if we lowered the bottom wage by ε. Since all workers are fully employed, it follows that all firms will choose to be representative to avoid the hiring penalty, and that $\tilde{w}(0) = w^{CE}(0) = w^*$ and $\tilde{w}(h)$ is increasing in h. Finally, since the marginal penalty for a deviation in a type's net wage from the economy-wide lowest type's wage is given by

$$-C \left[\tilde{w}(h) - \tilde{w}(0) \right]^2,$$

and since this cost goes to infinity as $C \to \infty$ for any positive wage gap, it follows that as C becomes large $\tilde{w}(h) \to \tilde{w}(0)$, and all of the workers are paid as if they were the lowest health status type and all of their productivity gap is absorbed by the cost of discriminating on wages. *Q.E.D.* ■

The fact that the productivity advantage of higher health status individuals is completely absorbed by the discrimination costs means that the society as a whole gets no gain from their productivity advantage. So the health expenditures that raise their productivity above the lowest type are inefficient. In addition, expenditure on the lowest health type relaxes the wage discrimination penalty on other types. So this equilibrium outcome is not socially efficient.

D Wages in the Competitive Equilibrium

To understand the implications of Proposition 6 for the behavior of equilibrium wages, note that our results imply that the equilibrium competitive wage is given by

$$\begin{aligned} w^{CE}(h) &= g(h)F(h,0) + (1-g(h))\int_0^{\bar{\varepsilon}^{CE}(h)} f(\varepsilon)F(h,\varepsilon - x(\varepsilon,h))d\varepsilon \\ & \quad +(1-g(h))\int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F(h,\bar{\varepsilon}^{CE}(h))d\varepsilon. \end{aligned}$$

Hence

$$\begin{aligned} \frac{dw^{CE}(h)}{dh} &= g'(h)\left[\begin{array}{c} F(h,0) - \int_0^{\bar{\varepsilon}^{CE}(h)} f(\varepsilon)F(h,\varepsilon - x(\varepsilon,h))d\varepsilon \\ -\int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F(h,\bar{\varepsilon}^{CE}(h))d\varepsilon \end{array}\right] \\ & \quad +g(h)F_1(h,0) + (1-g(h))\int_0^{\bar{\varepsilon}^{CE}(h)} f(\varepsilon)F_1(h,\varepsilon - x(\varepsilon,h))d\varepsilon \\ & \quad +(1-g(h))\int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F_1(h,\bar{\varepsilon}^{CE}(h))d\varepsilon \\ & \quad +(1-g(h))\int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F_2(h,\bar{\varepsilon}^{CE}(h))\frac{d\bar{\varepsilon}^{CE}(h)}{dh}d\varepsilon, \end{aligned}$$

since the net effect of the change in the integrand bounds generated by $\frac{d\bar{\varepsilon}^{CE}(h)}{dh}$ is zero. Next note that our optimality condition for $\bar{\varepsilon}^{CE}(h)$, (8), implies that

$$F_{12}(h,\bar{\varepsilon}^{CE}(h))dh + F_{22}(h,\bar{\varepsilon}^{CE}(h))d\bar{\varepsilon}^{CE}(h) = 0,$$

and hence

$$\frac{d\bar{\varepsilon}^{CE}(h)}{dh} = \frac{-F_{12}(h,\bar{\varepsilon}^{CE}(h))}{F_{22}(h,\bar{\varepsilon}^{CE}(h))}.$$

This result, along with (8), implies that

$$\begin{aligned} \frac{dw^{CE}(h)}{dh} &= g'(h)\left[\begin{array}{c} F(h,0) - \int_0^{\bar{\varepsilon}^{CE}(h)} f(\varepsilon)F(h,\varepsilon - x(\varepsilon,h))d\varepsilon \\ -\int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F(h,\bar{\varepsilon}^{CE}(h))d\varepsilon \end{array}\right] \\ & \quad +g(h)F_1(h,0) + (1-g(h))\int_0^{\bar{\varepsilon}^{CE}(h)} f(\varepsilon)F_1(h,\varepsilon - x(\varepsilon,h))d\varepsilon \\ & \quad +(1-g(h))\int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F_1(h,\bar{\varepsilon}^{CE}(h))d\varepsilon \\ & \quad -(1-g(h))\int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F_2(h,\bar{\varepsilon}^{CE}(h))\frac{F_{12}(h,\bar{\varepsilon}^{CE}(h))}{F_{22}(h,\bar{\varepsilon}^{CE}(h))}d\varepsilon. \end{aligned} \tag{37}$$

All of the terms in (37) are trivially positive except the last, which is negative since $F_{22} < 0$. However, so long as the spillover ratio F_{12}/F_{22} evaluated at $(h,\bar{\varepsilon}^{CE}(h))$ is not too negative, then wages will vary positively with health status. Note that this is trivially implied if the direct effect of the change in health status offsets the spillover, or

$$F_1(h, \bar{\varepsilon}^{CE}(h)) - F_2(h, \bar{\varepsilon}^{CE}(h)) \frac{F_{12}(h, \bar{\varepsilon}^{CE}(h))}{F_{22}(h, \bar{\varepsilon}^{CE}(h))} > 0. \tag{38}$$

Note that this is a condition purely on the fundamentals of the economy since $\bar{\varepsilon}^{CE}(h)$ is given by an (implicit) equation that depends only on exogenous elements in the model. We summarize our results in the following proposition:

Proposition 15 *The competitive wage is increasing in h if (37) is positive.*

Now we show that, instead of assuming that production firms offer a combination of a wage and a health insurance contract to the worker, we could alternatively have assumed that wage contracts are offered by competitive production firms and health insurance contracts are offered by *separate* competitive health insurers, without changing the equilibrium allocation of health expenditures and consumption characterized in the main text.

To achieve this, in the next proposition we construct a wage contract offered by production firms such that it is in the worker's interest to buy the competitive health insurance contract characterized in Proposition 6. Furthermore, the production firm's payoff under this wage contract is independent of whether a worker has in fact bought health insurance. As a consequence, the allocation from Proposition 6 can be obtained without bundling wage and health insurance contracts, and without the need for the production firms to be able to verify whether and what health insurance workers have chosen to buy. All the firm has to observe in order to implement this wage contract is the worker's current productivity.

Proposition 16 *A wage contract of the form*

$$\boldsymbol{w}(h, \varepsilon - x) = \begin{cases} w^{CE}(h) & \text{if } F(h, \varepsilon - x) \geq F(h, \bar{\varepsilon}^{CE}(h)) \\ w^{CE}(h) - \left[F(h, \bar{\varepsilon}^{CE}(h)) - F(h, \varepsilon - x)\right] & \text{if } F(h, \varepsilon - x) < F(h, \bar{\varepsilon}^{CE}(h)) \end{cases} \tag{39}$$

offered by production firms implements the allocation characterized in Proposition 6 as a competitive equilibrium in which households purchase health insurance contracts of the form $\{x^{CE}(\varepsilon, h), P^{CE}(h)\}$ from competitive health insurers.

The proof is in Appendix B along with the proofs of all other theoretical results. Note that in this decentralization the production firm implements partial consumption insurance against the ε health shocks (those below $\bar{\varepsilon}^{CE}(h)$), with the health insurance contract providing the remainder.[65] Also note that an individual who faces this wage contract but cannot buy health insurance will still find it optimal to individually make the competitive health insurance expenditures $x^{CE}(\varepsilon, h)$

E Computation of the Social Planner's Problem

The social planner's problem in section 4.2.1 can be solved numerically, either by making the problem recursive or by brute force optimization over the finite-dimensional vectors $\{c_t(h), e_t(h), V_t(h)\}$. The planner's recursive problem has as a state variable the cross-sectional distribution over health status Φ, which makes it rather cumbersome to solve. Instead, we solve the sequence problem directly, using a penalty function approach to ensure that the aggregate resource constraint is satisfied in every period t. Thus the problem

[65] One interpretation of the contract is that firms have limited commitment and thus can only provide partial ε insurance.

we solve is

$$\max_{\{c_t(h),e_t(h),V_t(h)\}_{t=0}^T} \sum_h \Phi_0(h)V_0(h) - \sum_{t=1}^T P\left(Y(\Phi_t) - \sum_h c_t(h)\Phi_t(h)\right)$$

$$s.t.$$

$$V_t(h) = u(c_t(h)) - q(e_t(h)) + \beta \sum_{h'} Q(h'; h, e_t(h))v_{t+1}(h') \quad (40)$$

$$q'(e_t(h)) = \beta \sum_{h'} \frac{\partial Q(h'; h, e_t(h))}{\partial e_t(h)} v_{t+1}(h'), \quad (41)$$

$$\Phi_{t+1}(h') = \sum_h Q(h'; h, e_t(h))\Phi_t(h). \quad (42)$$

where the penalty function P is given by

$$P(x) = \frac{\kappa}{2}(\min\{0, x\})^2 = \begin{cases} 0 & \text{if } x \geq 0 \\ \frac{\kappa}{2}(x)^2 & \text{if } x < 0 \end{cases}$$

where κ is a penalty parameter. Ideally we want κ to be large (to make sure the constraints are satisfied at the optimal solution), but the larger κ is, the harder it might be to solve the optimization problem. This suggests the following algorithm,

Algorithm 17 *Choose κ^0 small. Solve the above maximization problem with κ^0, and denote the solution as $\left\{c_t^0(h), e_t^0(h), V_t^0(h)\right\}_{t=1}^T$ and denote the solution in iteration step n as $\{c_t^n(h), e_t^n(h), V_t^n(h)\}_{t=0}^T$. Then iterate on*

1. *For given κ^n solve the maximization problem using $\left\{c_t^{n-1}(h), e_t^{n-1}(h), V_t^{n-1}(h)\right\}_{t=0}^T$ as an initial guess. The optimal solution is $\{c_t^n(h), e_t^n(h), V_t^n(h)\}_{t=0}^T$*

2. *Update*

$$\kappa^{n+1} = \phi\kappa^n$$

where ϕ is a computational parameter that trades off speed (high ϕ) and stability (low ϕ).

3. *Stop if*

$$\left\|c^{n+1} - c^n\right\| < \varepsilon.$$

F Computation of the Equilibrium with a No Prior Conditions Law and/or a No Wage Discrimination Law

The algorithm to solve this version of the model shares its basic features with that for the social planner's problem, but differs in terms of the sequence of variables on which we iterate. We describe in detail the algorithm for the no prior conditions law and then briefly discuss how it is modified for the other policy regimes:

Algorithm 18 1. *Guess a sequence $\{Eu'_t, P_t\}_{t=0}^T$.*

2. *Given the guess use equations* (9) *and* (3) *to determine health cutoffs and wages $\{\bar{\varepsilon}_t^{NP}(h), w_t(h)\}$.*

3. *Given $\{w_t(h), P_t\}$, solve the household's dynamic programming problem* (15) *for a sequence of optimal effort policies $\{e_t(h)\}_{t=0}^T$.*

4. *From the initial health distribution Φ_0 use the effort functions $\{e_t(h)\}_{t=0}^T$ to derive the sequence of health distributions $\{\Phi_t\}_{t=0}^T$ from equation* (11).

5. *Obtain a new sequence $\{Eu_t'^{new}, P_t^{new}\}_{t=0}^T$ from* (18) *and* (19).

6. If $\{Eu_t'^{new}, P_t^{new}\}_{t=0}^T = \{Eu_t', P_t\}_{t=0}^T$ we are done. If not, go to Step 1. with new guess $\{Eu_t'^{new}, P_t^{new}\}_{t=0}^T$.

Note that instead of the $Eu_t' = \sum_h u'(w^{NP}(h;\Phi_t) - P^{NP}(\Phi_t))\Phi_t(h)$ one could iterate on $\{w_t(h)\}$ which is more transparent, but significantly increases the dimensionality of the problem.

The algorithm for no wage discrimination is a slight modification of that for no prior conditions. The algorithm iterates over $\{Eu_t', w_t\}_{t=0}^T$. In Step 2 given the guess use equations (10) and (5) to determine health cutoffs and premia $\{\bar{\varepsilon}_t^{NP}(h), P_t(h)\}$. In Step 4 obtain a new sequence $\{Eu_t'^{new}, w_t^{new}\}_{t=0}^T$ from (17) and (20). With both policies, equation (20) replaces (19) in all expressions.

G Details for Data and Estimation

G.1 Augmented Model Analysis: Inclusion of the z-shock

We assume that households *must* incur the cost z, when the z-shock hits. This assumption and the fact that households are risk averse imply that the z-shock will be fully insured in the competitive equilibrium under any policy (and of course by the social planner). Moreover, we assume that households receiving a z-shock can still work at full productivity. Therefore, in a competitive equilibrium, the wage of a worker with health status h is unchanged, and the health insurance premium is determined as

$$P(h) = (1 - g(h)) \int_0^{\bar{\varepsilon}} x(\varepsilon, h) f(\varepsilon) d\varepsilon + \mu_z(h).$$

Given our assumptions there is no interaction between the z-shocks and the health insurance contract problem associated with the ε-shock, since it is prohibitively costly by assumption not to bear the z-expenditures. The role of the z-expenditures is to soak up the most extreme health expenditures observed in the data associated with catastrophic illnesses, but to otherwise leave our theory from the previous sections unaffected.

The static analysis goes through completely unchanged in the presence of the z-shocks. In the dynamic analysis the benefits of higher effort e and thus a better health distribution $\Phi_t(h)$ now also include a lower mean catastrophic health expenditure $\mu_z(h)$. This extension of the model leads to straightforward extensions of the expressions derived in the analysis of the dynamic model in section 4.2, and does not change any of the theoretical properties derived in sections 4.1 and 4.2.

G.2 Descriptive Statistics

Before we proceed to descriptive statistics of the PSID data, we summarize, in Table 12, the mapping between variables in our model and data.

Table 12: Mapping between Data and Model

Model	Description	Data	
		Variable	Corresponding Years
x, μ_z	Medical Expenditure	Average total expenditure reported in 1997-2002 in MEPS	1997-2002
w	Labor Income	Average total labor income reported in 1999, 2001, 2003 in PSID	1998,2000,2002
h	Health status	1997 in MEPS and PSID	1997
e	Effort	Average effort measures reported in 1999, 2001, and 2003 in PSID	1998,2000,2002

Since our model period is six years, we take averages of reported medical expenditure and labor income over the six-year periods that we observe. Moreover, we use health status data from 1997 (rather than 1999) to capture the effect of health on wages and medical expenditure.

Table 13 documents descriptive statistics of key variables that we use in our analysis. All data are in 2000 dollars. The reported statistics are from the PSID survey years 1999, 2001, and 2003, except for medical

Table 13: Descriptive Statistics

	Mean	St. Dev.	Median
Age	43.879	10.941	44
Health Status	2.774	0.941	3
Labor Income	40,684	51,287	32,280
fair health	15,616	18,631	9,324
good health	34,270	33,506	28,720
very good health	43,433	47,731	34,971
excellent health	53,838	72,336	40,350
Medical Expenditure	2,262	5,623	783
fair health	5,675	11,220	2,248
good health	2,407	5,124	931
very good health	1,780	4,430	732
excellent health	1,322	2,483	515
Light Physical Exercise (e.g. walking, bowling)	0.616	0.290	0.666
Heavy Physical Exercise (e.g. jogging, swimming)	0.265	0.260	0.190
Smoking (number of cigarettes per day)	4.373	8.809	0.000

expenditures. The reported labor income is the average labor income over three survey years. Medical expenditure data are from the MEPS and are averaged over six survey years (1997 through 2002). While the PSID only reports medical expenditures at the household level, the MEPS contains individual-level medical expenditure, which is the relevant statistic for us. In both surveys, health status is reported as Excellent, Very Good, Good, Fair, and Poor, but we group Poor with Fair to ensure a large enough sample size in the bottom health group (poor health status constitutes less than 5% of the sample). Excellent health status is assigned a numerical value of 4, in a descending order with Fair, which has a value of 1.

G.3 Estimation of the Health Transition Function

The PSID reports information about an individual's exercise level starting in 1999. We use all available data between 1999 and 2013 to estimate the health transition function. We chose a model period of six years, and therefore, we use six-year changes in health status for our estimation. Table 14 summarizes the transition matrix for health status from the raw, untreated PSID data.

Table 14: Health Transition over 6 years

	Fair	Good	Very Good	Excellent
Fair	0.622	0.262	0.089	0.027
Good	0.195	0.495	0.252	0.058
Very Good	0.065	0.302	0.511	0.122
Excellent	0.033	0.133	0.358	0.476

Using the functional form described in the main body of the paper, we estimate the health transition function, separately by each education group, in the following way. To recall, the set of parameters to be estimated for each education group is:

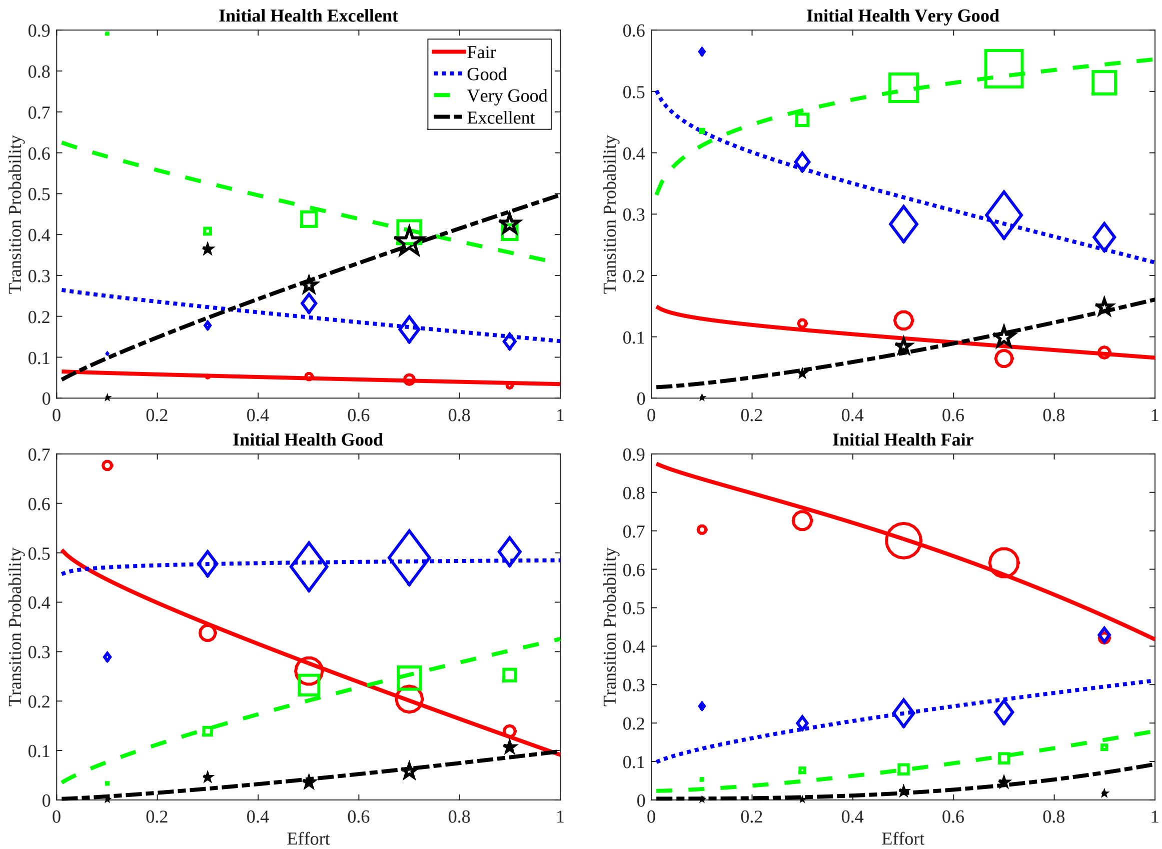
$$\boldsymbol{\theta} = \Big\{\{G(h,h')\}, \{\delta_l, \delta_h\, \delta_s\}, \phi(h), \lambda(h), \alpha_1(h), \alpha_2(h)\Big\}.$$

We use maximum likelihood estimation to estimate these parameters, where the log-likelihood is given by

$$\log \mathcal{L}(\boldsymbol{\theta}) = \sum_{obs=1}^{N_{obs}} \mathbf{1}\{h' = h'_{obs}\} log[Q(h'|h_{obs}, e^l_{obs}, e^h_{obs}, 1 - s_{obs})].$$

The estimated parameter values for each education group are summarized in Table 4 in the main text. Moreover, we plot the estimated transition function by each initial health group and *education* in Figures 16 and 17 (smooth lines). These figures complement Figure 4 in the main text, which is based on education-pooled data.

Figure 16: Transition for Low Education

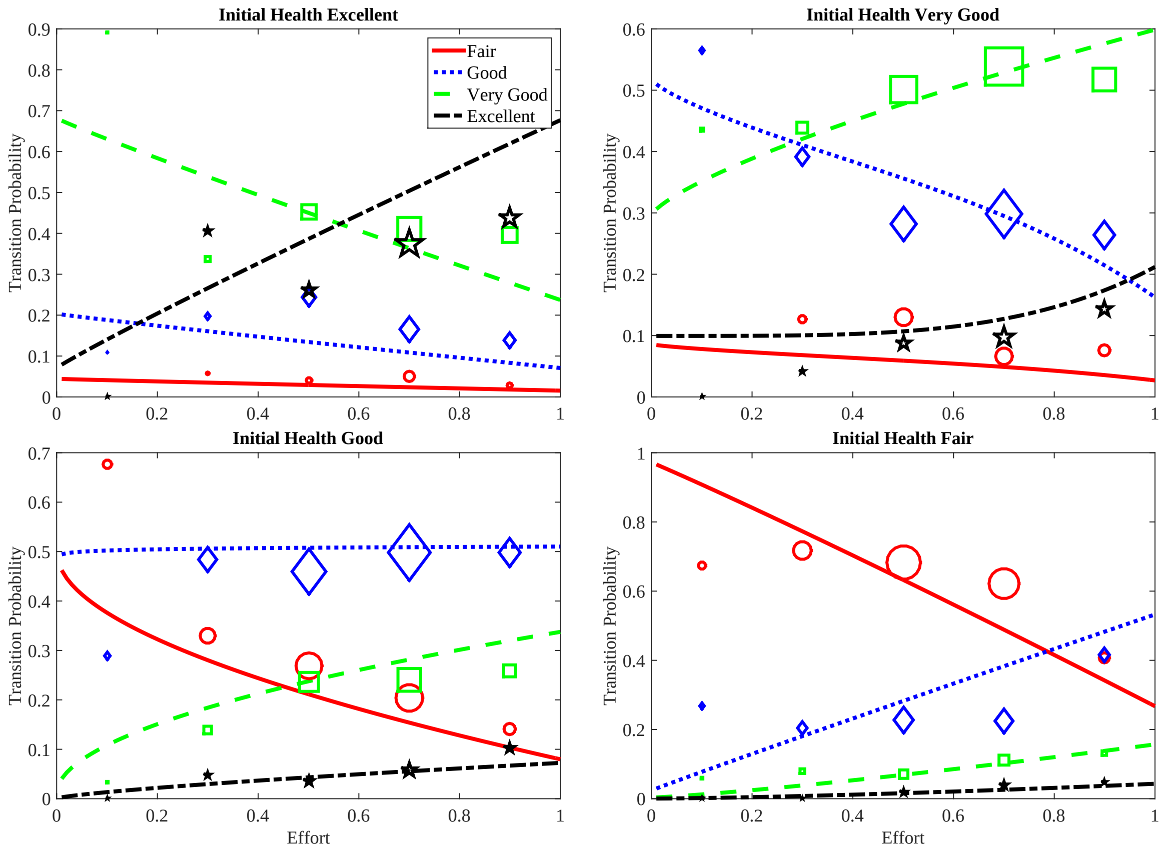


Note: Solid lines represent the estimated health transition functions for the low education group. The scatter plots are data points, with the size of markers representing the sample sizes.

Effects of Exercise on Health Status: Evidence from the Empirical Literature In this section, we investigate the plausibility of our estimates determining the effects of effort on health transition using empirical findings. One of the key difficulties lies in that most of the existing studies in the literature focus on a specific disease (e.g., diabetes, hypertension), whereas we use the self-reported health statuses. In the next, we first summarize the empirical findings from Colman and Dave (2012) and several medical journals. Then, we describe how we map our findings to theirs.

Colman and Dave use the first National Health and Nutrition Examination Survey (NHANES I) conducted between 1971 and 1974 and its follow-up study, the NHANES I Epidemiologic Follow-up Study (NHEFS) spanning 1982 to 1984 to estimate the impact of physical activity on the risk factors for heart disease. Using an exercise survey (reported as very active, moderately active, or quite inactive), they find that physical exercise reduces risk factors for diseases, and has a lagged effect that endures over time from the fixed-effects model (which controls for the unobserved time-invariant characteristics) and the lagged outcome model (which controls for time-varying factors). They find a significant decrease in the BMI between 0.28 and 0.78 points (1.2-3% relative to the mean) from a high level of exercises (either excluding recreational exercise or with only recreational exercise), relative to those in the lowest tercile in the exercise distribution.

Figure 17: Transition for High Education



Note: Solid lines represent the estimated health transition functions for the high education group. The scatter plots are data points, with the size of markers representing the sample sizes.

Moreover, they report that recreational exercise reduces the probability of having hypertension by between 2.8 and 8.4 percentage points (10-31% relative to the mean) depending on intensity, using the lagged outcome model. Similarly, the effect ranges between 1.8 and 2.2 percentage points (39-43%) for diabetes, and 1.9 and 2.4 percentage points (23-29%) for heart disease. They also report the effects of non-recreational exercises and estimates from the fixed-effects model.

We further summarize, in Table 15, some of the findings in the medical literature and their method of research.

One of the key difficulties lies in that most of the existing studies in the literature focus on a specific disease (e.g., diabetes, hypertension), whereas we use the self-reported health statuses. In order to map empirical findings to ours, we conduct the following analyses:

(a) Construct a mapping between health status and prevalence of disease (from the MEPS, Table 16).

(b) Using our estimated health transition function, for each initial health status, calculate a person's probability of having a disease in the future by exercise tercile (3 groups, consistent with Colman and Dave, 2012) and quintile (5 groups) using the formula:

$$Prob(disease'|h, e) = \sum_{h'} \left[Q(h'|h, e) \times Prob(disease|h') \right].$$

(c) Obtain the implied probability of having a health condition (through changes in health status) for

Table 15: Effects of Physical Activity on Disease: Empirical Evidence from the Medical Literature

Disease	Effects of Exercise	Method	Source
Diabetes	• Exercise (0.5 hr/day) and diet changes lower incidence by 58%	Randomized trial	Tuomilehto, et al. (2001)
	• Moderate exercise of 2.5 hr/week lowers incidence by 44%	Randomized trial	Hamman, et al (2006)
	• Runners' incidence lower by 41%	Observational study over 6 years	Williams & Thompson (2013)
Hypertension	• Runners' incidence lower by 14%	Observational study over 6 years	Williams & Thompson (2013)
	• Aerobic exercise lowers systolic and diastolic blood pressure by 3.8mm hg and 2.58 mm hg, respectively	Meta-analysis of randomized trials	Whelton, et al. (2002)
Coronary Heart Disease	• Running (1 hr/wk), weight training (0.5 hr/wk), and rowing (1 hr/wk)/walking (0.5 hr/day) lowers incidence by 42%, 23%, and 18% respectively	Observational study	Tanasescu, et al. (2002)
Cancer	• 30-60 min/day activity lowers colon and breast cancer incidence by 30-40% and 20-30%, respectively	Median measures from 50-60 epidemiologic studies	Lee (2003)

different exercise groups, and compare the magnitude to those in Colman and Dave (2012) and the medical literature (summarized in Table 15).

For the first step, mapping the prevalence of health conditions to health status, we use MEPS 2000 and 2001 data for all but cancer, for which we use 2008 data, the earliest year for which cancer diagnosis data are available. The result is presented in Table 16, and as expected, the prevalence of adverse health conditions is higher for those with fair health compared to those with excellent health.

Table 16: BMI and Adverse Health Condition Prevalence by Health Status

	BMI	Diabetes	Hypertension	Heart Disease	CHD	Cancer	Breast CA	Colon CA
Fair	29.4	0.2205	0.4897	0.2307	0.0708	0.1289	0.0148	0.0052
Good	28.4	0.0856	0.3040	0.0879	0.0204	0.0861	0.0148	0.0041
Very Good	27.0	0.0339	0.1927	0.0526	0.0094	0.0632	0.0090	0.0024
Excellent	26.0	0.0120	0.1094	0.0307	0.0023	0.0495	0.0053	0.0005
Total	27.4	0.0640	0.2377	0.0761	0.0173	0.0743	0.0102	0.0027

Note: This table documents the share of population with a specific health condition by health status. We use MEPS 2000 and 2001 data, except for cancer (any, breast, and colon), which is based on MEPS 2008, the earliest year in which the survey asks whether the individual has been diagnosed with cancer. The sample is restricted to those between the ages of 24 and 65. An individual is reported to have heart disease if he was diagnosed with coronary heart disease (CHD); angina or angina pectoris; heart attack or myocardial infarction; or another kind of heart disease or condition; or a stroke or transient ischemic attack.

Our results from the experiment following Colman and Dave (2012) are presented in Table 17. It documents the percentage change in the incidence of disease relative to the mean of an individual's initial health status (average over education). In the last row, we also document the findings by Colman and Dave. For example, our transition function predicts a drop of around 4 - 6% in the incidence of hypertension, while

Colman and Dave find about a 10-31% decline. For BMI and other diseases, we see that the estimated impact of effort on health is smaller under our health transition function than those in Colman and Dave's paper.

Table 17: Effect of Exercise (by Tercile) on BMI and the Incidence of Diseases from the Estimated Health Transition: Percent Change Relative to the Mean (within Health Status)

Disease	BMI			Diabetes			Hypertension			Heart Disease		
Exercise	Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
Fair	0.58	0.01	-0.59	9.68	0.24	-9.92	5.83	0.14	-5.97	9.24	0.23	-9.48
Good	0.49	-0.03	-0.46	10.16	-0.59	-9.57	5.36	-0.31	-5.05	8.92	-0.52	-8.40
Very Good	0.37	0.02	-0.39	7.22	0.28	-7.50	3.86	0.17	-4.03	4.89	0.18	-5.07
Excellent	0.40	0.02	-0.42	8.74	0.37	-9.11	5.05	0.22	-5.27	5.43	0.24	-5.67
Colman & Dave	1.2–3% decrease			39–43% decrease			10–31% decrease			23–29% decrease		

Note: This table documents the change in BMI, and the prevalence of diseases from the estimated transition function (using the procedure described in the text). The reported numbers are expressed as a percent change from the within-health status mean. For example, if a person with fair health status exerts effort in the lowest third among all fair people, he will experience an increase in his BMI of 0.58% relative to the mean (within fair health) BMI (or he is 5.83% more likely to experience hypertension), while if he had exerted the highest exercise, he would have seen a 0.59% drop in BMI (a 5.97% drop in hypertension).

Moreover, in Table 18, we document the prevalence of disease by initial health status and effort quintiles, normalized by the prevalence for the lowest effort quintile group. According to our estimated health transition function, individuals in the top quintile of exercise experience between a 20 and 24% lower incidence of diabetes, compared to those in the lowest quintile of exercise. In the randomized controlled trials, Tuomilehto, et al. (2001) and Hamman, et al (2006) find a 44-58% decrease in the incidence of diabetes, which is higher than our estimates. Similarly, our estimated effects for hypertension, coronary heart disease, and breast and colon cancer are similar to or lower than those documented in the medical literature, implying that our health transition functions do not over-estimate the effect of effort on health status over time.

G.4 Labor Income Profile

In this section and the next, we describe how we generate the labor income and medical expenditure profiles that we use as targets in estimating the model.

1. Run $\ln w_i = \beta_0 + \beta(t, educ, h) D_i(t, educ, h) + \beta_z Z_i$, where
   - $D_i(t, e, h)$: Agebin ($t = 1, 2, 3, ..., 7$), education ($educ = 1, 2$) and health status ($h = 1, 2, 3, 4$) dummy for individual i
   - Z_i: Dummy variables for male and ethnicity
2. Use the coefficients to back out the joint effect of $(t, educ, h)$ on labor income. Let
$$\ln \tilde{w}(t = \hat{t}, e = \hat{e}, h = \hat{h}) = \hat{\beta}_0 + \hat{\beta}(t = \hat{t}, e = \hat{e}, h = \hat{h}) + C,$$
   where C ensures that the average of $\ln \tilde{w}$ is equal to the average of $\ln w$ from the data.
3. Smooth out the wage schedules by fitting them to a quadratic function in age:
$$\ln \tilde{w}(t, educ, h) = \gamma_0 + \gamma_1 t + \gamma_2 t^2,$$
   for each $(educ, h)$ group.

We use the smoothed wage profiles as targets for the estimation, which are documented in Table 22.

Table 18: Effect of Exercise on Disease Prevalence from the Estimated Health Transition

Disease	Health	Effort Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Diabetes	Fair	1.00	0.92	0.89	0.86	0.77
	Good	1.00	0.90	0.87	0.83	0.76
	Very Good	1.00	0.94	0.91	0.88	0.82
	Excellent	1.00	0.94	0.91	0.87	0.80
Hypertension	Fair	1.00	0.95	0.93	0.91	0.85
	Good	1.00	0.94	0.93	0.91	0.87
	Very Good	1.00	0.97	0.95	0.93	0.90
	Excellent	1.00	0.96	0.94	0.92	0.88
Coronary Heart Disease	Fair	1.00	0.91	0.87	0.84	0.74
	Good	1.00	0.88	0.85	0.80	0.73
	Very Good	1.00	0.94	0.91	0.88	0.82
	Excellent	1.00	0.93	0.90	0.85	0.78
Breast Cancer	Fair	1.00	0.98	0.98	0.97	0.95
	Good	1.00	0.97	0.96	0.95	0.93
	Very Good	1.00	0.97	0.96	0.94	0.91
	Excellent	1.00	0.97	0.95	0.92	0.89
Colon Cancer	Fair	1.00	0.96	0.95	0.93	0.89
	Good	1.00	0.95	0.94	0.92	0.89
	Very Good	1.00	0.96	0.94	0.92	0.87
	Excellent	1.00	0.94	0.90	0.85	0.78

Note: Effects of exercise on incidence of diseases are calculated from our estimated transition function (using the process described in the text). We normalize the disease incidence of the lowest effort quintile within health status to 1. For example, those with excellent health whose exercise level is in the top 20% (within their health status) are expected to have a 20% lower probability (0.80) of having diabetes, compared to those whose exercise level is in the bottom 20%.

Effects of Health on Income: Evidence from Disease Prevalence As discussed in the main text, Bartel and Taubman (1979) find significant negative effects of health conditions (e.g., heart disease, hypertension, arthritis, bronchitis, disease of the liver, gallbladder and pancreas) that result in reductions in earnings ranging between 20 and 30%. Focusing on arthritis and controlling for selection bias, Mitchell and Butler (1986) estimate a 33% decline, and Mitchell and Burkhauser (1990) estimate between a 19 and 32% decline in wage income. People with diabetes and complications are also found to experience earnings losses between 28 and 33%, according to Kraut, et al. (2001) and Ng, et al. (2001).

Normalizing the income of workers without a disease to 1, we calculate the expected earnings of workers of health status h as

$$y(disease) \times Prob(disease|h) + 1 \times (1 - Prob(disease|h)).$$

In Table 19, we document the prevalence of hypertension and arthritis and implied earnings by health status, using a 25% income loss for hypertension (a mid-point from Mitchell and Butler), and 30% for arthritis and diabetes. We find that they lead to a loss of about 11 to 17% in expected earnings, lower than but still roughly similar to the income ratios used in the robustness analysis with a lower health-income gradient (which we document in Section I.1), especially for young workers.

It is also worth noting that people with lower health statuses tend to have multiple health conditions. Among workers with fair health statuses, 32% have two or more health conditions reported, and 11% of those with good health have two or more conditions.

Table 19: Disease Incidence and Implied Earnings Ratio by Health Status[1]

	Hypertension		Arthritis		Diabetes	
	Prevalence	Ratio	Prevalence	Ratio	Prevalence	Ratio
Fair	0.490	0.902	0.522	0.871	0.221	0.937
Good	0.304	0.950	0.293	0.942	0.086	0.978
Very Good	0.193	0.979	0.120	0.972	0.034	0.993
Excellent	0.109	1.000	0.105	1.000	0.012	1.000

[1] This is based on MEPS 2000 and 2001 data. Sample is restricted to those between the ages of 24 and 65.

G.5 Health Shocks and Medical Expenditure Profile

For the quantitative analysis, we model catastrophic medical expenditure shocks, which we define as medical expenditures in excess of \$10,000. Only about 5% of the population spend more than \$10,000 in medical expenditure, and given the lowest income level of \$15,000, it seems reasonable to consider the shock catastrophic.

As we described in the main body of the paper, the expected expenditure from catastrophic shocks, $\mu_z(t, educ, h)$ is exogenously given. We construct these parameters using procedures similar to those for labor income. First, we construct the smoothed total medical expenditure and medical expenditure on ε-shocks, using expenditures less than \$10,000. Then, we subtract the two to obtain $\mu_z(t, educ, h)$ and they are documented in Table 20.

Table 20: Expected Catastrophic Medical Expenditures

	HS Graduates				Some College			
	Fair	Good	Very Good	Excellent	Fair	Good	Very Good	Excellent
Age 24-29	356	228	159	140	523	364	255	186
Age 30-35	508	268	177	161	864	402	302	204
Age 36-41	709	334	213	190	1,328	476	362	239
Age 42-47	964	440	276	231	1,949	602	442	299
Age 48-53	1,275	606	382	288	2,769	810	547	397
Age 54-59	1,641	868	563	369	3,840	1,146	687	555
Age 60-65	2,051	1,281	879	488	5,227	1,693	875	814

For our ε-shock expenditure profiles, we again use the same procedure as for labor income, and use the smoothed moments as our targets, which are presented in Table 22.

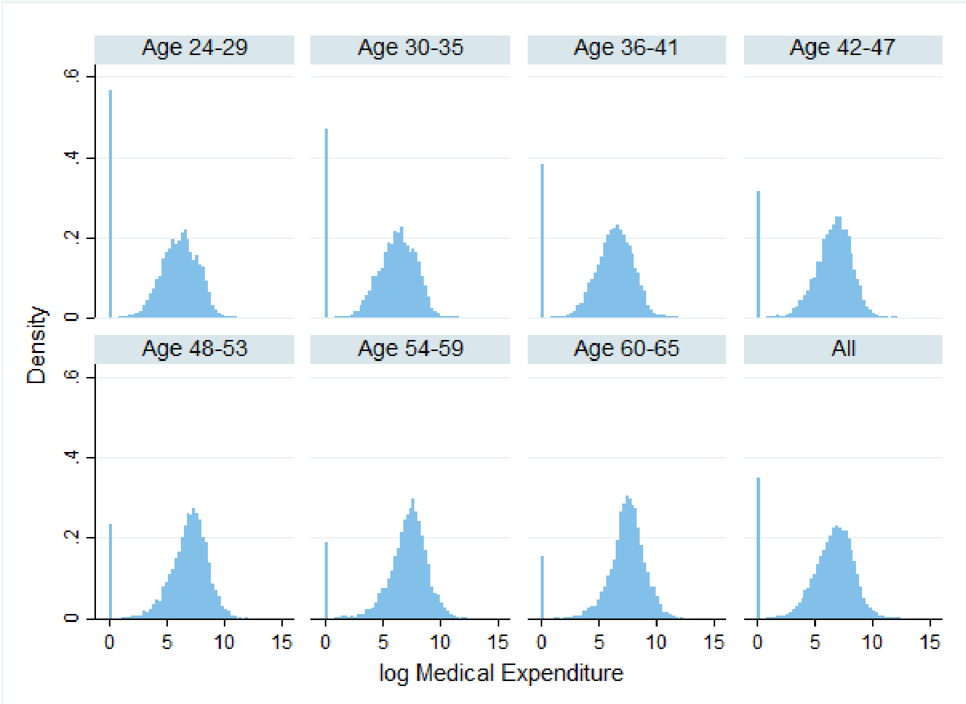
The remaining parameters for the health shock distribution are the health-dependent probabilities of not getting health shocks $\tilde{g}(h)$, the parameter that governs age-effect on probability α_g, the mean and variance of ε-shock distribution μ_ε and σ_ε^2, and two parameters to reflect the age-effect on mean and variance α_μ and α_σ. In our model, medical expenditures on ε-shocks are endogenously determined by productivity concerns, and we estimate these parameters within the model.

Figure 18 plots medical expenditure distributions by age group, and for all. It is clear that there are shifts in medical expenditure distributions over time, and our parameterizations are aimed at capturing such shifts in the health shock distributions, and consequently in medical expenditures.

G.6 Estimation Results

Table 21 summarizes the value of parameters that were estimated inside the model using efficient GMM. Table 22 contains the details of the model fit with respect to income and medical expenditure profiles.

Figure 18: Distribution of (log) Medical Expenditure



H Additional Quantitative Results

H.1 Welfare Comparisons by Education Group

Tables 23 presents the static and dynamic consumption equivalent variations for each education group.

H.2 Limited Wage Insurance

In section 6.4, we document aggregate welfare effects when the no wage discrimination law is not perfectly enforced ($\tau > 0$). In Figure 19, we plot average effort levels and the share of population with excellent or very good health when $\tau = 0.2$ (the optimal τ from the social welfare perspective), against our benchmark case. As is evident from the plots, with partial incentives through the labor market, the reduction in health effort is more modest, e.g., approximately 11% during the main working ages, for the no wage discrimination policy, relative to the competitive equilibrium (and smaller for individuals toward retirement). The resulting deterioration of the health distribution is also smaller than under our benchmark case.

I Other Robustness and Sensitivity Analysis: Details

I.1 Lower Health-Income Gradient

In order to assess the robustness of our results with respect to the magnitude of the health-income gradient, we re-estimate the model to match income profiles with a lower income gradient of health and evaluate the policy consequences. In Figure 20, we plot the new income profiles used as empirical targets. We halve the health-income gradient with respect to Good health, in order to keep the average income level in the economy the same as in the benchmark estimation. The earnings ratio implied by our benchmark estimation and our lower gradient estimation for Fair (normalized to 1), Good, Very Good, and Excellent health are

Table 21: Parameters Estimated within the Model

Parameters	Value	Parameters	Value
h_1	6,214	$\alpha_t(t=1)$	1.068
h_2	9,589	$\alpha_t(t=2)$	1.109
h_3	10,423	$\alpha_t(t=3)$	1.134
h_4	11,548	$\alpha_t(t=4)$	1.152
$A(t=1, educ=1, \tilde{h}=1)$	2.184	$\alpha_t(t=5)$	1.156
$A(t=2, educ=1, \tilde{h}=1)$	1.792	$\alpha_t(t=6)$	1.151
$A(t=3, educ=1, \tilde{h}=1)$	1.439	$\alpha_t(t=7)$	1.141
$A(t=4, educ=1, \tilde{h}=1)$	1.202	$\alpha_e(educ=2)$	1.032
$A(t=5, educ=1, \tilde{h}=1)$	1.029	$\phi(educ=1)$	9.708
$A(t=6, educ=1, \tilde{h}=1)$	0.954	$\phi(educ=2)$	9.638
$A(t=7, educ=1, \tilde{h}=1)$	0.800	$\xi(educ=1)$	1.837
$A(t=1, educ=2, \tilde{h}=1)$	1.523	$\xi(educ=2)$	1.860
$A(t=2, educ=2, \tilde{h}=1)$	1.265	$g(h=1)$	0.114
$A(t=3, educ=2, \tilde{h}=1)$	1.115	$g(h=2)$	0.144
$A(t=4, educ=2, \tilde{h}=1)$	0.973	$g(h=3)$	0.144
$A(t=5, educ=2, \tilde{h}=1)$	0.883	$g(h=4)$	0.155
$A(t=6, educ=2, \tilde{h}=1)$	0.818	α_g	0.206
$A(t=7, educ=2, \tilde{h}=1)$	0.789	α_μ	0.138
$A(t=1, educ=1, \tilde{h}=2)$	2.071	α_σ	0.010
$A(t=2, educ=1, \tilde{h}=2)$	1.856	μ_ε	1,922
$A(t=3, educ=1, \tilde{h}=2)$	1.661	σ^2_ε	564,550
$A(t=4, educ=1, \tilde{h}=2)$	1.256	$\gamma(h=1)$	0.029
$A(t=5, educ=1, \tilde{h}=2)$	1.047	$\gamma(h=2)$	0.019
$A(t=6, educ=1, \tilde{h}=2)$	0.781	$\gamma(h=3)$	0.003
$A(t=7, educ=1, \tilde{h}=2)$	0.539	$\gamma(h=4)$	0.002
$A(t=1, educ=2, \tilde{h}=2)$	1.740	ψ	3.163
$A(t=2, educ=2, \tilde{h}=2)$	1.379	$v_{T+1}(educ=1, h=2)$	0.042
$A(t=3, educ=2, \tilde{h}=2)$	1.177	$v_{T+1}(educ=1, h=3)$	0.149
$A(t=4, educ=2, \tilde{h}=2)$	1.025	$v_{T+1}(educ=1, h=4)$	0.201
$A(t=5, educ=2, \tilde{h}=2)$	0.933	$v_{T+1}(educ=2, h=2)$	0.120
$A(t=6, educ=2, \tilde{h}=2)$	0.847	$v_{T+1}(educ=2, h=3)$	0.198
$A(t=7, educ=2, \tilde{h}=2)$	0.757	$v_{T+1}(educ=2, h=4)$	0.458

respectively,

$$\begin{aligned}
\text{Benchmark estimation, All:} \quad & [1.00; 1.42; 1.53; 1.75] \\
\text{Benchmark estimation, Young (Old):} \quad & [1.00(1.00); 1.16(1.61); 1.29(1.72); 1.41(2.01)] \\
\text{Lower gradient estimation, All:} \quad & [1.00; 1.16; 1.21; 1.29] \\
\text{Lower gradient estimation, Young (Old):} \quad & [1.00(1.00); 1.07(1.22); 1.13(1.26); 1.19(1.37)],
\end{aligned}$$

where young are between the ages of 24 and 41, and old, between the ages of 42 and 65. As is clear from these numbers, the overall substantial health-income gradient is primarily driven by the health-income differences among older workers.

As we discussed in section 5.2.3 of the main paper and section G.4 in the appendix, these income gradients, especially those used for our robustness analysis, are within the ranges found in the empirical literature.

Table 22: Model Fit on Labor Income and Medical Expenditures

	Labor Income				Medical Expenditure			
	Low Education		High Education		Low Education		High Education	
	Model	Data	Model	Data	Model	Data	Model	Data
Age 24-29, fair health	24,180	26,781	31,480	32,124	1,944	1,405	4,456	5,160
good health	26,616	27,395	41,973	41,566	1,179	1,058	2,369	2,665
very good health	29,134	29,632	46,078	43,329	1,012	941	2,071	2,108
excellent	32,556	29,136	51,661	49,880	857	784	1,734	1,545
Age 30-35, fair health	28,337	28,587	39,481	40,779	2,151	2,405	3,918	3,780
good health	32,083	31,323	48,756	48,567	1,602	1,703	2,205	2,244
very good health	35,233	35,487	53,704	53,536	1,423	1,322	1,770	1,772
excellent	39,529	37,509	60,459	60,282	1,249	1,081	1,388	1,419
Age 36-41, fair health	28,852	28,838	45,868	44,684	2,310	1,876	5,608	6,605
good health	35,203	34,103	53,712	52,754	1,280	1,230	2,900	3,299
very good health	38,751	39,188	59,308	60,251	1,064	1,017	2,482	2,522
excellent	43,604	43,877	66,967	67,023	879	872	2,113	1,886
Age 42-47, fair health	27,983	27,493	40,453	42,262	2,764	3,114	4,764	4,517
good health	36,165	35,356	54,867	53,269	1,737	1,915	2,838	2,882
very good health	39,869	39,903	60,677	61,760	1,543	1,526	2,298	2,373
excellent	44,941	46,636	68,642	68,552	1,310	1,170	1,781	1,763
Age 48-53, fair health	24,887	24,770	34,412	34,503	2,764	2,436	6,926	8,421
good health	34,858	34,904	52,127	50,003	1,524	1,466	3,625	4,217
very good health	38,439	37,465	57,670	57,663	1,251	1,160	2,990	3,054
excellent	43,340	45,041	65,271	64,504	1,023	998	2,612	2,396
Age 54-59, fair health	21,864	21,090	24,097	24,314	3,580	4,016	5,555	5,249
good health	29,714	32,811	44,790	43,634	2,016	2,223	3,789	3,795
very good health	32,751	32,436	49,532	49,036	1,787	1,783	3,140	3,352
excellent	36,904	39,526	56,031	55,837	1,500	1,317	2,406	2,254
Age 60-65, fair health	16,377	16,970	14,803	14,789	3,292	3,077	8,330	10,695
good health	25,463	29,370	35,603	35,396	1,784	1,791	4,568	5,565
very good health	28,036	25,893	39,331	37,982	1,436	1,396	3,549	3,743
excellent	31,547	31,518	44,436	44,465	1,144	1,173	3,206	3,171

Table 23: Welfare Comparisons by Education

	Static CEV^i		Dynamic CEV^i	
	HS Graduates	Some College	HS Graduates	Some College
Constrained Social Planner	1.017	1.452	3.301	7.442
Competitive Equilibrium	0.000	0.000	0.000	0.000
No Prior Conditions Law	0.219	0.170	1.534	4.018
No Wage Discrimination Law	1.009	1.450	2.780	6.907
Both Policies	1.017	1.452	0.346	5.112

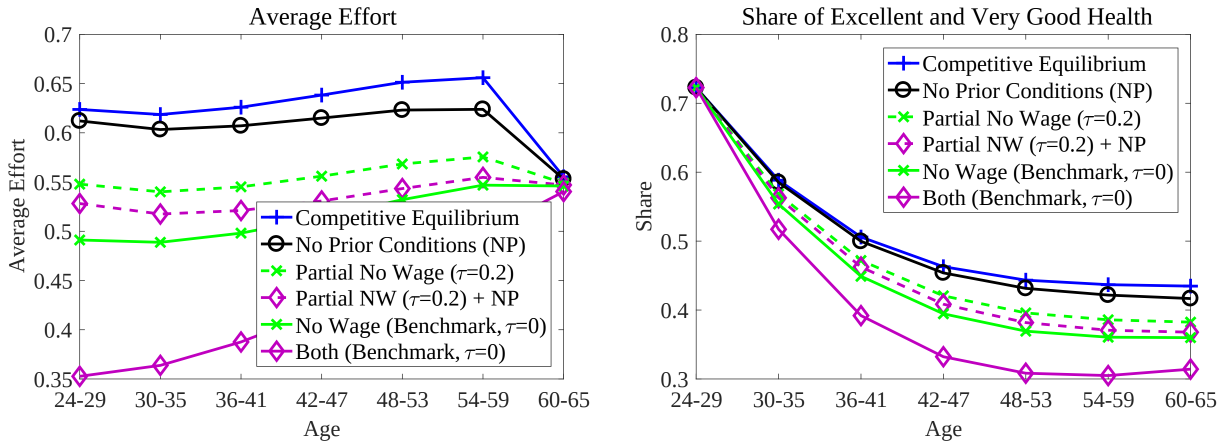
I.2 Endogenous Labor Supply

Now assume that preferences are given by

$$u(h) = v\left(c(h) - \lambda\frac{l(h)^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}\right)$$

where v is a strictly concave and differentiable function. We could assume that v is of CRRA form and do so for the quantitative analysis, but this assumption is not necessary for the theoretical arguments to follow.

Figure 19: Effort and Health Effects from a Partial No Wage Discrimination Policy



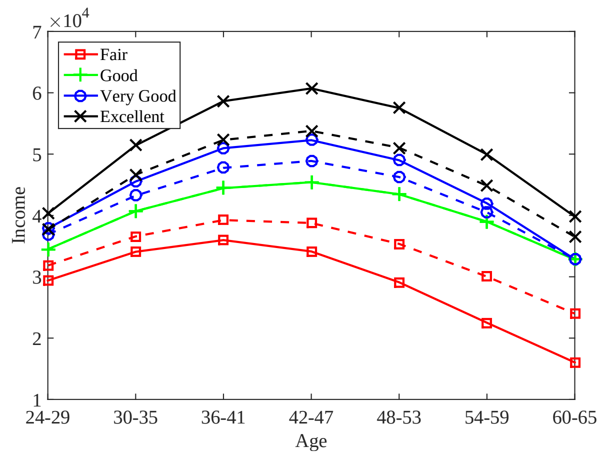


Figure 20: Labor Income Profiles: Benchmark (Solid) vs. Lower Health-Income Gradient (Dotted)

Output per unit of labor is given by $w(h, \varepsilon) = F(h, \varepsilon - x)$. In the model with exogenous income we assumed that the firm can insure the worker against ε shocks, and we continue to make this assumption here.

I.2.1 Competitive Equilibrium Without and With Policy

In the competitive equilibrium, firms provide wage insurance against ε, and hours and health expenditures are chosen jointly as part of the optimal contract. The problem reads

$$U^i = \max_{x(\varepsilon,h),w(h),P(h),l(h)} \sum_h \Phi(h) v \left(w(h)l(h) - P(h) - \lambda \frac{l(h)^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}} \right)$$

subject to

$$w^i(h) = g(h)F(h,0) + (1-g(h))\int_0^1 f(\varepsilon)F(h,\varepsilon - x(\varepsilon,h))d\varepsilon, \quad i \in \{CE, NP\} \tag{43}$$

$$w = \sum_h \Phi(h)\left[g(h)F(h,0) + (1-g(h))\int_0^1 f(\varepsilon)F(h,\varepsilon - x(\varepsilon,h))d\varepsilon\right], \quad i \in \{NW, B\} \tag{44}$$

$$P^i(h) = (1-g(h))\int_0^1 f(\varepsilon)x(\varepsilon,h)d\varepsilon, \quad i \in \{CE, NW\} \tag{45}$$

$$P = \sum_h \Phi(h)\left[(1-g(h))\int_0^1 f(\varepsilon)x(\varepsilon,h)d\varepsilon\right], \quad i \in \{NP, B\} \tag{46}$$

with associated wage and premium functions under different policies, $i \in \{CE, NP, NW, B\}$.

Competitive Equilibrium Let the Lagrange multipliers be μ_w and μ_P. The first-order conditions with respect to $l(h), w(h), P(h), x(\varepsilon, h)$ are given by

$$\begin{aligned} v'(.)\left(w(h) - \lambda l(h)^{\frac{1}{\chi}}\right) &= 0 \\ v'(.)l(h) &= \mu_w \\ -v'(.) &= \mu_P \\ \mu_w(1-g(h))f(\varepsilon)F_2(h,\varepsilon - x(\varepsilon,h)) &= \mu_P(1-g(h))f(\varepsilon). \end{aligned}$$

Thus, combining the equations, optimal labor supply and health expenditure cutoffs are determined by

$$l(h) = [w(h)/\lambda]^{\chi} \tag{47}$$

$$l(h)F_2(h,\varepsilon^{CE}(h)) = -1, \tag{48}$$

and wages and premia are obtained from equations (43) and (45), evaluated with $x(\varepsilon,h) = \max[0, \varepsilon - \bar{\varepsilon}^{CE}(h)]$.

Equations (47) and (48) are static efficiency conditions with endogenous labor supply. First, in the absence of income effects, optimal labor supply is increasing in wage (health status). Second, the optimal health insurance contracts are now adjusted to account for the fact that healthy workers work more. We discuss the impact of the latter on health expenditures further in the following section.

Denote by income $y(h) = w(h)l(h) = w(h)^{1+\chi}\lambda^{-\chi}$, then we can write static utility as

$$u^{CE}(h) = v\left(w(h)l(h) - P(h) - \lambda\frac{l(h)^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}\right) = v\left(\frac{y(h)}{1+\chi} - P(h)\right).$$

Thus, conditional on the new health insurance contract, the consequences of endogenous labor supply is to scale down the income-health gradient by a factor $\frac{1}{1+\chi}$. This feature also holds under the policy equilibria we discuss below.

No Prior Conditions Under no prior conditions, the first-order conditions read as

$$\begin{aligned} l(h) &= (w(h)/\lambda)^{\chi} \\ l(h)F_2(h,\bar{\varepsilon}^{NP}(h)) &= -\frac{\sum_h \Phi(\tilde{h})v'\left(\frac{w(\tilde{h})^{1+\chi}\lambda^{-\chi}}{1+\chi} - P\right)}{v'\left(\frac{w(h)^{1+\chi}\lambda^{-\chi}}{1+\chi} - P\right)} \end{aligned}$$

Thus, as was the case in the benchmark model, the cutoff of a health type h is determined as the ratio between average marginal utilities and his own marginal utility, adjusted for labor supply. Wages and premium are determined by equations (43) and (46), respectively, with $\bar{\varepsilon}^{NP}(h)$.

No Wage Discrimination Similarly, the first-order conditions read

$$
\begin{aligned}
l(h) &= l = [w/\lambda]^{\chi} \\
l(h)F_2(h,\bar{\varepsilon}^{NW}(h)) &= -\frac{v'\left(wl(h) - P(h) - \lambda(h)\frac{l(h)^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}\right)}{\sum_h \Phi(\tilde{h})v'\left(wl(\tilde{h}) - P(\tilde{h}) - \lambda(\tilde{h})\frac{l(\tilde{h})^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}\right)}.
\end{aligned}
$$

With a constant wage rate for all, the labor supply of workers is also equalized across health statuses. Wages and premium are determined by equations (44) and (45), respectively, for given health expenditure cutoffs $\bar{\varepsilon}^{NW}(h)$.

Both Policies From the first-order conditions we immediately find

$$
\begin{aligned}
l(h) &= l = [w/\lambda]^{\chi} \\
F_2(h,\bar{\varepsilon}^{B}(h)) &= -\left[\frac{\lambda}{w}\right]^{\chi}.
\end{aligned}
$$

Moreover, wages and premium are determined by equations (44) and (46), evaluated at $\bar{\varepsilon}^{B}(h)$. Given a constant wage rate, the optimal labor supply chosen by workers is constant in health status h. Thus, under both policies, there is full utility insurance against health status risks. However, as we will see from the social planner's allocation, this is not the efficient way of providing full utility insurance.

I.2.2 Social Planner's Problem

To provide the most relevant comparison, we also assume that the social planner is restricted to not condition hours on ε. The planner chooses hours and expenditures optimally, and the problem becomes

$$
\begin{aligned}
U^{SP} &= \max_{c(\varepsilon,h),x(\varepsilon,h),l(h)} \sum_h \Phi(h)\left(g(h)v\left(c(0,h) - \lambda\frac{l(h)^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}\right) + (1-g(h))\int f(\varepsilon)v\left(c(\varepsilon,h) - \lambda\frac{l(h)^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}\right)d\varepsilon\right) \\
s.t. \quad & \sum_h \Phi(h)\left[g(h)\left[c(0,h)+x(0,h)\right] + (1-g(h))\int f(\varepsilon)[c(\varepsilon,h)+x(\varepsilon,h)]d\varepsilon\right] \\
&= \sum_h \Phi(h)\left[g(h)l(h)F(h,-x(0,h)) + (1-g(h))l(h)\int_0^1 f(\varepsilon)F(h,\varepsilon - x(\varepsilon,h))d\varepsilon\right]
\end{aligned}
$$

The first-order condition with respect to consumption is (with μ being the Lagrange multiplier on the resource constraint):

$$
v'\left(c(\varepsilon,h) - \lambda\frac{l(h)^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}\right) = \mu
$$

which immediately implies $c(\varepsilon,h) = c(h)$ for all h.

With respect to labor we have

$$
\begin{aligned}
& -\lambda l(h)^{\frac{1}{\chi}}\Phi(h)\left(g(h)v'\left(c(0,h) - \lambda\frac{l(h)^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}\right) + (1-g(h))\int f(\varepsilon)v'\left(c(\varepsilon,h) - \lambda\frac{l(h)^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}\right)d\varepsilon\right) \\
= \; & -\mu\Phi(h)\left[g(h)F(h,-x(0,h)) + (1-g(h))\int_0^1 f(\varepsilon)F(h,\varepsilon - x(\varepsilon,h))d\varepsilon\right]
\end{aligned}
$$

and simplifying

$$l^{SP}(h) = \left[\frac{w^{SP}(h)}{\lambda}\right]^{\chi}$$

where

$$w^{SP}(h) = \left[g(h)F(h, -x(0,h)) + (1-g(h))\int_0^1 f(\varepsilon)F(h, \varepsilon - x(\varepsilon, h))d\varepsilon\right].$$

The efficient allocation of health expenditure is given by

$$-l^{SP}(h)F_2(h, \bar{\varepsilon}^{SP}(h)) = 1$$

and thus cutoffs and "wages" for the social planner are given by, for all h:

$$\begin{aligned} F_2(h, \bar{\varepsilon}^{SP}(h)) &= -\left[\frac{\lambda}{w^{SP}(h)}\right]^{\chi} \\ w^{SP}(h) &= \left[g(h)F(h,0) + (1-g(h))\int_0^1 f(\varepsilon)F(h, \varepsilon - x^{SP}(\varepsilon, h))d\varepsilon\right] \end{aligned}$$

Incidentally, these thresholds are exactly identical to those of the competitive equilibrium. Thus the competitive equilibrium health thresholds continue to be efficient even with endogenous labor supply. Total net output is then determined by $Y = \sum_h \Phi(h)\left[l(h)w^{SP}(h)\right]$, and is the same as in the competitive equilibrium since health expenditure thresholds, productivities and hours are the same as in the competitive equilibrium.

It remains to determine the efficient allocation of consumption. As with an exogenous labor supply, the inefficiency is imperfect utility insurance. The social planner finds it optimal to equate

$$c(h) - \lambda\frac{l(h)^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}$$

across all h, and given labor supply $l(h) = \left[\frac{w^{SP}(h)}{\lambda}\right]^{\chi}$ chooses the consumption levels $c(h)$ such that $c(h) - \lambda\frac{l(h)^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}$ are equated across h. Since $l(h)$ is increasing in h, so is $c(h)$. The consumption levels solve

$$\begin{aligned} \sum_h \Phi(h)c(h) &= Y \\ c(h) - \lambda\frac{l(h)^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}} &= U \text{ for all } h \end{aligned}$$

these are $H+1$ equations in the unknowns $c(h)$ and U. The two basic insights from the original static problem carry over unchanged: *(a)* the competitive health insurance contract is efficient and *(b)* the social planner provides additional h insurance. Now the planner fully insures the utility kernel $c - \frac{l^{1+\frac{1}{\chi}}}{1+\frac{1}{\chi}}$ rather than just c and thus finds it optimal to have c and l co-vary positively with h, but the basic source of inefficiency is the same.

It is instructive to compare the solution to the static social planner's problem and the static outcome under both policies. With an exogenous labor supply, the solutions to both problems were identical as both had the same efficient health expenditure schedule and full consumption insurance. The latter characteristic is maintained with an endogenous labor supply, as in both problems the planner and the policies provide perfect *period utility insurance,* and thus the combination of both policies is fully effective in providing social insurance. However, the social planner can achieve this by efficiently having labor supply depend on health ($l^{SP}(h)$ that is increasing in h) and then compensate high h individuals with higher consumption so that $c^{SP}(h)$ is also increasing in h. In the competitive equilibrium the only way to provide this insurance is through full wage and premium insurance, with the unintended consequence that labor supply is inefficiently uncorrelated with productivity. Also, and indirectly, the lack of a spread in labor supply also makes the health insurance allocation inefficient (since the planner finds it optimal, ceteris paribus, to give more generous health insurance to those working longer hours). Thus, we would not expect the combination of both policies, even statically, to reproduce the efficient allocation.

I.2.3 Characterization of Optimal Health Insurance Contract

Now we analyze the impact of an endogenous labor supply on the health insurance contract. We focus on the competitive equilibrium and the combination of both policies, since contrasting these two cases gives the strongest and clearest distortions of the health insurance contract from the social insurance policies.

Competitive Equilibrium To make the results more intuitive, let $\eta = \bar{h}$ denote the mean health level, and assume that λ is calibrated in such a way that $l(\bar{h}) = 1$. Then $\bar{\varepsilon}^{CE}(\bar{h})$ remains the same as in the benchmark with an exogenous labor supply. Furthermore

$$\frac{F_2(h, \bar{\varepsilon}^{CE}(h))}{F_2(h, \bar{\varepsilon}^{CE}(\bar{h}))} = \left[\frac{w(\bar{h})}{w(h)}\right]^{\chi}.$$

In the benchmark $\chi = 0$ and thus the right hand side is 1. For $\chi > 0$ and conjecturing that it continues to be true that $w(h)$ is strictly increasing in h, then $\frac{w(\bar{h})}{w(h)} < 1$ for $h > \bar{h}$ and $\frac{w(\bar{h})}{w(h)} > 1$ for $h < \bar{h}$. Thus, relative to the benchmark model $\bar{\varepsilon}^{CE}(h)$ is less increasing in h (and might in fact be decreasing). That is, relative to the benchmark, health insurance is relatively more generous for healthier individuals and relatively less generous for sicker (lower h) individuals. This in turn will spread out wages across health levels, in turn accentuating income differences across health levels. That is, relative to the $\chi = 0$ case the health-income gradient $y(h)$ itself is likely to be steeper. It is important to note that because of the more generous health insurance contract the spread in $P(h)$ will be reduced as well (which in turn reduces the insurance benefit of the no prior conditions policy).

Both Policies Since the adoption of both policies leads to the most extreme deterioration of incentives relative to the competitive equilibrium, we contrast the previous result with the impact of an endogenous labor supply under both policies. In this policy configuration the thresholds are determined

$$F_2(h, \bar{\varepsilon}^{CE}(h)) = -\left[\frac{\lambda}{w}\right]^{\chi}$$

and thus

$$\frac{F_2(h, \bar{\varepsilon}^{CE}(h))}{F_2(h, \bar{\varepsilon}^{CE}(\bar{h}))} = 1$$

as in the benchmark with an exogenous labor supply. Thus, the relative health shock thresholds remain unchanged with both policies in place. This is in contrast to the spreading out of the health-income gradient in the competitive equilibrium due to the endogenous adjustment of the health insurance contract in response to labor supply being endogenous.

I.2.4 Summary and Interpretation

In summary, an endogenous labor supply changes two elements, one that directly affects the trade-off between incentives and insurance, and can be directly assessed by shrinking the health-income gradient by a factor $1 + \chi$, for a fixed health expenditure levels. Moreover, there is an interaction of an endogenous labor supply with the optimal health insurance contract, since healthier individuals work longer hours (at least in the absence of income effects; their presence will make this argument less quantitatively relevant and might even overturn it). Thus they should receive more generous health insurance (relative to the exogenous labor supply). As the previous section shows, this second, indirect effect will tend to spread the health gradient in the competitive equilibrium out again, relative to the exogenous labor benchmark.

Our quantitative results show that with an endogenous labor supply, households have another margin of smoothing static utility across different health levels h. This is turn reduces the welfare effects of the policies compared to the benchmark model.

I.3 Insurance Pooling as the Benchmark

Table 24 summarizes the key parameter changes induced by switching the estimated economy to the partial no prior conditions economy.

Table 24: Parameter Values with Insurance Pooling (PL) as the Benchmark

	CE	PL		CE	PL
$\gamma(h=1)$	0.029	0.016	$v_{T+1}(educ=1,h=2)$	0.042	0.042
$\gamma(h=2)$	0.019	0.010	$v_{T+1}(educ=1,h=3)$	0.149	0.147
$\gamma(h=3)$	0.003	0.002	$v_{T+1}(educ=1,h=4)$	0.201	0.199
$\gamma(h=4)$	0.002	0.001	$v_{T+1}(educ=2,h=2)$	0.120	0.116
ψ	3.163	3.541	$v_{T+1}(educ=2,h=3)$	0.198	0.198
			$v_{T+1}(educ=2,h=4)$	0.458	0.454

I.4 Mismeasured Effort Inputs

For a given household, let the parameter vector be given by $\chi = (\rho, \nu)$, and assume that the cost function is of the form:

$$C(\tilde{e}, e^*, \chi) = \Psi(\chi)\left[\rho\tilde{e}^{\nu} + (1-\rho)(e^*)^{\nu}\right]^{\frac{1}{\nu}}.$$

Maximizing $C(\tilde{e}, e^*, \chi)$ subject to equation (24) yields as the optimality condition

$$\frac{\tilde{e}}{e^*} = \left[\frac{\lambda(1-\rho)}{(1-\lambda)\rho}\right]^{\frac{1}{\nu-1}} := \kappa(\rho, \nu; \lambda). \tag{49}$$

Using this equation in 24 yields as the optimal solution $\tilde{e} = \Gamma_1(\rho, \nu; \lambda)e$ and $e^* = \Gamma_2(\rho, \nu, \lambda)e$, where

$$\Gamma_1(\rho, \nu; \lambda) = \frac{\kappa(\rho,\nu;\lambda)}{\lambda\kappa(\rho,\nu;\lambda)+1-\lambda} \tag{50}$$

$$\Gamma_2(\rho, \nu; \lambda) = \frac{1}{\lambda\kappa(\rho,\nu;\lambda)+1-\lambda} \tag{51}$$

Plugging these solutions into 24 it follows that

$$\begin{aligned} e &= \lambda\tilde{e} + (1-\lambda)\frac{\Gamma_2(\rho,\nu,\lambda)}{\Gamma_1(\rho,\nu,\lambda)}\tilde{e} \\ &= \eta\lambda\tilde{e}, \end{aligned}$$

where, for a fixed λ, the random variable $\eta = 1 + \frac{(1-\lambda)\Gamma_2(\rho,\nu,\lambda)}{\lambda\Gamma_1(\rho,\nu,\lambda)}$ has a cross-sectional distribution determined by the population distribution $F(\rho, \nu)$. The function $\Psi(\chi)$ can be chosen such that $C(\tilde{e}(e), e^*(e), \chi) = e$ and thus $q(e)$ retains the interpretation as the utility cost of providing true effort e. This requires

$$\Psi(\chi)\left[\rho\Gamma_1^{\nu} + (1-\rho)\Gamma_2^{\nu}\right]^{\frac{1}{\nu}} = 1. \tag{52}$$

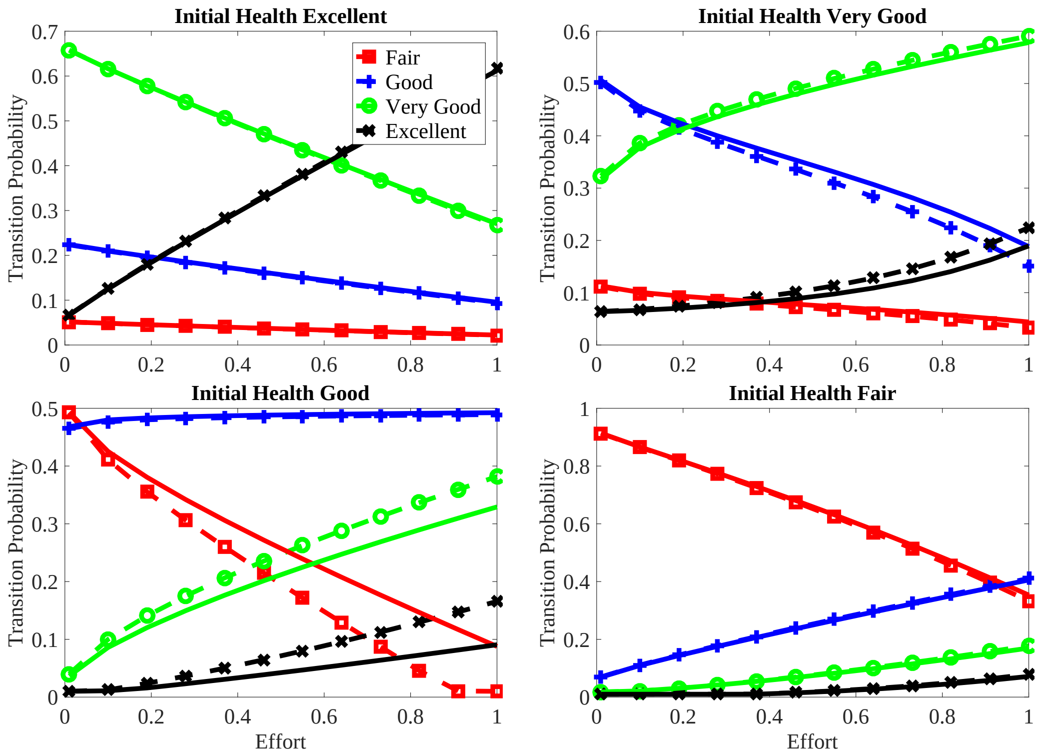
Note that under the assumptions on the cost function C, given the observation $\tilde{e}$ and given an assumed distribution for χ, the other aspects of the model are not needed to infer the distribution of e^*. This continues to permit us to estimate Q outside the model despite the fact that $\tilde{e}, e^*$ are endogenous variables. Second, also note that the distribution of e^* is not truncated for any possible observation of $\tilde{e}$ despite the fact that both $e, \tilde{e}$ are restricted to lie in the unit interval. Third, if the distribution of χ and thus of η is degenerate, then $\tilde{e}$ and e^* are perfectly correlated and $\tilde{e}$ becomes a perfect proxy for e. Hence, the extent of the measurement error with respect to e given $\tilde{e}$ depends upon the variance of χ and the size of λ. Finally, note that the lower $\tilde{e}$ is, the smaller is the support of $e = \eta\lambda\tilde{e}$. This implies that high $\tilde{e}$ observations are subject to more doubt than low $\tilde{e}$ observations. Thus, the measurement error is not classical, since $E(e|\tilde{e}) = E\left\{1 + (1-\lambda)\eta\right\}\lambda\tilde{e} = \lambda\tilde{e}$.

We can therefore rewrite our maximum likelihood problem with noisy effort observations as

$$\max_{\theta} \sum_i \int_{2\lambda-1}^{1} \log\left[\frac{Q(h_i'|h_i, \eta\lambda\tilde{e}_i, \theta)}{(2-2\lambda)}\right] d\eta \tag{53}$$

where i indexes the household effort and health status observations. For our application in the main text, we set $\lambda = 0.75$. Figure 21 displays our new estimates of Q (weighted averages across education groups) against the estimated transitions in the benchmark.

Figure 21: Transition Functions with Measurement Error



Note: Solid lines represent our benchmark specification without measurement error. The dotted lines with markers represent the estimated transition functions with measurement error and $\lambda = 0.75$. The plots are weighted averages over education groups.

References

[1] Bartel A. and P. Taubman (1979), "Health and Labor Market Success: The Role of Various Diseases." *Review of Economics and Statistics*, 61(1), 1-8.

[2] Colman, G. and D. Dave (2012), "Physical Activity and Health," *International Journal of Arts & Sciences*, 5(6), 29-45.

[3] Colombo, Francesca and Nicole Tapay (2004), "Private Health Insurance in OECD Countries: The Benefits and Costs for Individuals and Health Systems," *OECD Health Working Papers No. 15.*

[4] European Commission (2012), "Supported employment for people with disabilities in the EU and EFTA-EEA."

[5] Hamman, RF, RR Wing, SL Edelstein, JM Lachin, GA Bray, L. Delahanty, M. Hoskin, AM Kriska, EJ Mayer-Davis, X. Pi-Sunyer, J. Regensteiner, B. Venditti, and J. Wylie-Rosett (2006), "Effect of weight loss with lifestyle intervention on risk of diabetes," *Diabetes Care*, 29(9), 2102-7.

[6] Hasegawa, Tamako (2010), "Japan's Employment Measures for Persons with Disabilities: Centered on Quota System of "Act on Employment Promotion of Persons with Disabilities," *Japan Labor Review*, 7(2).

[7] Kraut, Allen, Robert Tate, Randy Walld, and Cam Mustard (2001), "Impact of Diabetes on Employment and Income in Manitoba, Canada," *Diabetes Care*, 24:64-68.

[8] Lee, I-Min (2003), "Physical Activity and Cancer Prevention – Data from Epidemiologic Studies," *Medicine & Science in Sports & Exercise*, 35(11), 1823-1827.

[9] Mitchell, J. and R.V. Burkhauser (1990). "Disentangling the Effect of Arthritis on Earnings: A Simultaneous Estimate of Wage Rates and Hours Worked," *Applied Economics*, 22(10), 1291-1309.

[10] Mitchell, J. and J.S. Butler (1986), "Arthritis and the Earnings of Men: An Analysis Incorporating Selection Bias," *Journal of Health Economics*, 5(1), 81-98.

[11] Ng, Ying Chu, Philip Jacobs, and J.A. Johnson (2001), "Productivity Losses Associated with Diabetes in the U.S.," *Diabetes Care*, 24, 257-261.

[12] OECD (2010), "Sickness, Disability and Work: Breaking the Barriers."

[13] Tanasescu, Mihela, Michael F. Leitzmann, Eric B. Rimm, Walter C. Willett, Meir J. Stampfer, and Frank B. Hu (2002), "Exercise Type and Intensity in Relation to Coronary Heart Disease in Men," *Journal of the American Medical Association*, 288(16), 1994-2000.

[14] Tuomilehto J., J. Lindstrm, JG Eriksson, TT Valle, H. Hmlinen, P. Ilanne-Parikka, S. Keinnen-Kiukaanniemi, M. Laakso, A. Louheranta, M. Rastas, V. Salminen, M. Uusitupa; Finnish Diabetes Prevention Study Group (2001), "Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance," *The New England Journal of Medicine*, 344(18), 1343-50.

[15] Whelton, Seamus P., Ashley Chin, Xue Xin, and Jiang He (2002), "Effect of Aerobic Exercise on Blood Pressure: A Meta-Analysis of Randomized, Controlled Trials," *Annals of Internal Medicine*, 136, 493-503.

[16] Williams, Paul T. and Paul Thompson (2013), "Walking vs running for hypertension, cholesterol, & diabetes risk reduction," *Arteriosclerosis, Thrombosis, and Vascular Biology*, 33(5), 1085-1091.