BECKER FRIEDMAN INSTITUTE
FOR RESEARCH IN ECONOMICS
THE UNIVERSITY OF CHICAGO

Working Paper Series
Health Economics Series No. 2016-06

Analyzing the Effects of Insuring Health Risks

HAROLD L. COLE, SOOJIN KIM, AND DIRK KRUEGER

September 1, 2016

Keywords: Health Risks, Social Insurance, Health Effort Choices

JEL Codes: E61, H31, I18

**Becker Friedman Institute
for Research in Economics**

Contact:
773.702.5599
bfi@uchicago.edu
bfi.uchicago.edu

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

September 29, 2016

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. Our study is motivated by recent US legislation that has tightened regulations on wage discrimination against workers with poorer health status (such as the 2008 amendment of the Americans with Disability Act from 1990, the ADAAA) and that prohibits health insurance companies from charging different premiums for workers of different health (a provision in the Patient Protection and Affordable Care Act, PPACA, that went in effect in 2014). To do so we construct and estimate (using PSID and MEPS data) 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. 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 although the competitive equilibrium features too little consumption insurance and a combination of both policies is effective in providing such insurance period by period, it is suboptimal (from an ex-ante welfare perspective) to introduce both policies *jointly* since such a policy innovation severely undermines the incentives to lead healthier lives and thus induces a more rapid deterioration of the cohort health distribution over time. This effect more than offsets the static gains from better consumption insurance so that expected discounted lifetime utility is lower under both policies, relative to implementing one policy in isolation.

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 labor and health care 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 incentive of households to care about their health. The size of health related risks and the importance of potential incentive effects rests on three empirical observations, documented in section 2.1. First, a better health status increases a worker's productivity and thus labor earnings to a sizeable 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 and over time this can have an appreciable impact on the health distribution of the population.

The public policies we study in the paper are restricted, stylized versions of the 1990 American with Disability Act (ADA) and its 2008 amendment (the ADAAA), as well as the Affordable Care Act (PPACA)

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

of 2010.[1] We will show that these policies can provide additional social insurance against health-related labor income risk and medical expenditure risk. We will also show that these policies adversely impact the dynamic incentives to lead healthy lives. The purpose of the paper is to qualitatively and quantitatively evaluate this trade-off. We examine whether these policies alone or in combination improve social welfare. We want to clearly state from the outset that ours is *not* a comprehensive study of all aspects and provisions of the ADAAA and the ACA; rather we focus on the role of these policies in providing social insurance against health-related income and health insurance premium risk, as well as their negative incentive effects.

To analyze the impact of these policies we construct a dynamic life cycle model with endogenous, stochastically evolving health status. As in Grossman (1972), health status is an individual state variable that determines both a household's productivity at work and the likelihood that she will be subject to an adverse health-related shock. Such a shock in turn can be offset by medical expenditures against which individuals can purchase health insurance.[2] (as in Dey and Flinn 2005). Health status itself is persistent and changes stochastically over time, and its evolution is affected by the household's efforts to maintain their health, such as exercising and abstention from smoking. Since health-sustaining effort is a choice and health-related social insurance policies reduce households' economic incentives to maintain their health, a trade-off between the provision of social insurance and private incentives emerges, rendering the adoption of these policies a nontrivial policy design question. Furthermore, by endogenizing the demand for medical expenditures and private health insurance policies, private insurance contracts respond to the public policy regime.

We use the model as a theoretical and quantitative laboratory for the study of wage-based and the health-insurance based social insurance policies. We call the first policy a wage non-discrimination policy that prevents employers from paying workers differentially by health status (and argue in section 2.2.1 that it provides a good approximation to the actual U.S. policy environment under ADAAA). The second policy, which we name a no-prior conditions policy (and argue in section 2.2.2 captures key elements of the PPACA legislation), restricts health insurance companies from charging a different premium to workers with different initial health status. Our objective is to evaluate the impact of these policies on the evolution of the distributions of health status, earnings and health insurance costs, and, eventually, on social welfare.

In order to isolate the social insurance benefits and incentive costs of both policies we study the static and the dynamic impact of these policies. 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 productivity of the workforce.

To empirically implement our quantitative analysis we first estimate and calibrate the model using PSID and MEPS data to match key statistics on labor earnings, medical expenditures and observed physical exercise levels. We then use the parameterized version of the model as a laboratory to evaluate the consequences of the different policy options. Our results show that combination of wage non-discrimination law and no prior conditions law provides full insurance against health risks and implements the first-best consumption insurance allocation in the short run (that is, in the static model), but leads to a severe deterioration of incentives and thus the population health distribution in the long run (that is, in the dynamic model). This effect more than offsets the static gains from better consumption insurance so that expected discounted lifetime utility is lower under both policies, relative to implementing wage nondiscrimination legislation *alone*. Both policy options are strongly welfare improving relative to the competitive equilibrium, however.

In section 2 we set out the empirical facts justifying our modelling approach that will be used when mapping the model to the data, discuss the actual U.S. policies motivating our policy analysis and then relate our work to the existing literature. We describe the model and implementation of the two polices in the context of the model in section 3. The theoretical analysis of the static and dynamic version of the baseline model is contained in sections 4.1 and 4.2. 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 contains the main quantitative results of the paper with robustness analyses in 7 and 8 concludes. Proofs and details of the quantitative analysis are relegated to the appendix.

[1] Restricted in that we only evaluate selected aspects of both legislations, stylized because we relegate issues of costly implementation and enforcement to the sensitivity analysis in section 7.3.1.

[2] We also model catastrophic health shocks which require nondiscretionary health expenditures to avoid death.

2 Empirical Facts, Policies and Related Literature

2.1 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 households can affect the dynamic evolution of their health status by taking costly actions. The public policies under study then provide additional social insurance against labor income risk and health expenditure risk, but also impact adversely 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 household heterogeneity in observable and unobservable factors.

2.1.1 Health Status and Income

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.[3] 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.[4]

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

2.1.2 Health Status and Medical Expenditures

In table 2 we exploit data from the 1997 to 2002 waves of the Medical Expenditure Panel Survey (MEPS) to document the 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 MEPS for medical expenditures, as this data set reports individual-level medical expenditures, whereas PSID reports total medical expenditures only at the household level. Also, expenditures in 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 status: 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).

[3]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).

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

2.1.3 Effort and Health Status Updates

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 health status of individuals is impacted by their effort to lead a healthy life.[5]

The three columns 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 rows document that those individuals with effort above the cross-sectional average are more likely to retain or improve their health status. The differences in health dynamics across two effort groups are significant at a 5% confidence level (using a χ^2 test). The remaining four rows display that the positive association of effort and health transition persists if we control for the initial health status the individual starts from.[6]

Table 3: Effort and Health Dynamics over 6 years

Health	All		Bad Initial Health		Good Initial Health	
	Eff < Avg.	Eff ≥ Avg.	Eff < Avg.	Eff ≥ Avg.	Eff < Avg.	Eff ≥ 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

2.1.4 Summary of Facts and Motivation of Model Elements

The data presented in this section suggest a strong role for health status in the determination of income and medical expenditures, and an important role for individual effort in the dynamic updating of health status. We now build a model based on these three premises and then use it to evaluate, qualitatively and quantitatively, the trade-off between incentive costs and insurance benefits of social insurance policies in the labor and health insurance market. Prior to turning to the model we first briefly discuss, in the next subsection, the actual policy changes that motivate our model-based policy analysis, first focusing on policies that limit health-based wage dispersion and then turning to polices that prevent price discrimination in the health insurance market based on prior health conditions.

2.2 Institutional Background

We consider two sets of social insurance policies, insuring individuals against adverse income consequences of poor health in the labor market and high health insurance premia in the health insurance market. These policies are motivated by aspects of the American with Disabilities Act (ADA) and its amendment (ADAAA) in 2008, as well as provisions of the Patient Protection and Affordable Care Act (PPACA), or ACA for short. Before delving into the details of these legislations and our modelling of them, we want to stress from the outset that our objective is not to model all aspects of the ADAAA and the ACA, but rather only those elements of the law that provide social insurance against health-related income and insurance premium risk.

2.2.1 Income- and Employment Based Health Discrimination

In 1990 Congress enacted the Americans with Disabilities Act (ADA) to ensure that the disabled have equal access to employment opportunities. The ADA interprets a disability as an impairment that prevents or severely restricts an individual from performing activities that are of central importance to one's daily life. The ADA permits an employer to establish job-related qualifications on the basis of essential functions of the job. A core requirement of the ADA is the obligation of the employer to make a reasonable accommodation

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

[6]Bad health is defined as health status (Good, Fair, and Poor), and the good health group is composed of the top 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.

to qualified disabled people. These accommodations include acquiring or modifying equipments, making existing facilities accessible and usable, providing qualified readers and interpreters, as well as allowing for part-time or modified work schedules and restructuring of jobs.

The ADA Amendment Act (ADAAA) of 2008 rejected the narrow interpretation of the ADA, thereby broadening the notion of a disability. This included prohibiting the consideration of measures that reduce or mitigate the impact of a disability in determining whether someone is disabled. It also allowed people who are discriminated against on the basis of a perceived disability to pursue a claim on the basis of the ADA regardless of whether the disability limits, or is perceived to limit a major life activity. Under the ADAAA people can be disabled even if their disability is episodic or in remission. For example people whose cancer is in remission or whose diabetes is controlled by medication are still disabled under the act. Recently, conditions like morbid obesity and alcoholism have been deemed disabilities.[7]

The ADAAA mandates any worker who can perform the essential features of a job when granted a reasonable accommodation must share equally in all of the rewards of that job, and this explicitly includes compensation and promotions. Moreover, a worker cannot be charged for the cost of his accommodation.[8] Hence, the amended ADA can substantially compress the extent to which wages vary with health.

2.2.2 Insurance Cost and Exclusion Discrimination

The Health Insurance Portability and Accountability Act (HIPAA) of 1996 placed limits on the extent to which insurance companies could exclude people or deny coverage based upon pre-existing conditions. Insurers were allowed exclusion periods for coverage of pre-existing conditions, but these exclusion periods were reduced by the extent of prior insurance. Still, insurers were allowed to charge higher premiums based upon initial conditions, limit coverage and set lifetime limits on benefits.

The Patient Protection and Affordable Care Act of 2010 extended protection against pre-existing conditions. Beginning in 2010 children below the age of 19 could not be excluded from their parents' health insurance policy or denied treatment for pre-existing conditions. Beginning in 2014 this restriction now applies to adults as well. Moreover, insurance companies are no longer able to use health status to determine eligibility, benefits or premia. In addition, insurers are prevented from limiting lifetime or annual benefits.

2.2.3 Summary of Policies

It is our interpretation of the above legislative changes that, relative to 20 years ago, it is much more difficult now for employers to condition employment opportunities and compensation on the health status of their (potential) employees or to preferentially hire workers with better health. In addition, current legislation has made it increasingly difficult to condition the acceptance into, and insurance premia of health insurance plans on prior health conditions. The purpose of this paper is to analyze the aggregate and distributional consequences of these two legislative trends, with specific focus on their interactions.[9]

2.3 The Related Literature

Our paper contributes to the broader literature that examines the macroeconomic and distributional implications of health, health insurance and health care policy reform in dynamic models. These contributions include Grossman (1972), Ehrlich and Becker (1972), French and Jones (2011), Hall and Jones (2007), De Nardi, French, and Jones (2015), Jeske and Kitao (2009), Attanasio, Kitao and Violante (2011), Ales, Hosseini and Jones (2012), Halliday et al. (2016), Hansen, Hsu and Lee (2014), Kopecky and Koreshkova (2015), Ozkan (2014), and Pashchenko and Porapakkam (2013). From this literature, the most closely relaed work are Brügemann and Manovskii (2010) and Jung and Tran (2016) who also study the effects of PPACA on health insurance coverage and macroeconomic aggregates. These papers do not focus on the incentive effects

[7]See https://www.ada.gov/regs2016/final rule adaaa.html for the regulatory assessment by the Justice Department. See 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.

[8]See https://www.ada.gov/pubs/adastatute08.htm and https://www.eeoc.gov/eeoc/publications/ada18.cfm.

[9]Even under a strict interpretation of the amended Act, insurance is not complete for workers who take their accommodation in the form a leave or reduced hours. For this reason, we consider both a versions of the ADA with complete health-related income compression as well as versions with partial compressions.

on health efforts and thus health transitions induced by regulation in *both* the labor as well as the health insurance markets (and crucially, their interaction) that we formalize in our model. Similarly, Dey and Flinn (2005) focus on the impact of employer-sponsored health insurance on job mobility of workers, and in an extended model, Aizawa and Fang (2013) study the effects of the ACA. Neither paper is concerned with the interaction of social insurance policies in the labor and health sector and its combined effects on household health effort choice, however.

This paper builds on the empirical literature studying the impact of health or diseases on earnings and the determinants of the dynamics of health. The empirical literature studying the health-income nexus (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. Pijoan-Mas and Rios-Rull (2014) find an important dependence of health transition function on socio-economic status (most importantly education - dependence we also permit in our model). Moreover, many studies (e.g. Colman and Dave, 2013; Booth et al., 2012) find that individual behavior is a significant determinant in health status changes over time, the key premise of our paper, whereas medical expenditures have limited impact on long-run health outcomes (Baiker et al., 2013). There are also some evidence of the importance of incentives on health behavior and outcomes. Bhattacharya et al. (2011) use evidence from a Rand health insurance experiment to show that access to health insurance leads to increases in body mass and obesity, driven by the fact that insurance insulates people from the impact of their excess weight on their medical expenditure costs.[10]

Finally, related to our study of wage non-discrimination laws is the literature that studies the effect of the 1990 ADA legislation on employment, wages and labor hours of the disabled (see e.g. Acemoglu and Angrist, 2001; DeLeire, 2000; and DeLeire, 2001). Most find that it has decreased the employment of the disabled. DeLeire (2001) documents that the negative effects of poor health on the earnings of the disabled fell significantly in 1993 (after the implementation of the ADA), compared to the pre-reform year of 1984.

3 The Model

Time $t = 0, 1, 2, \dots T$ 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 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 $H = \{h_1, \dots, h_N\}$ 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 utility function u is twice differentiable, strictly increasing and strictly concave. 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.[11] In every period households with current health h remain healthy (that is, $\varepsilon = 0$) with probability $g(h)$. With probability $1 - g(h)$ the household draws a health shock $\varepsilon \in (0, \bar{\varepsilon}]$ which is distributed according to the probability density function $f(\varepsilon)$.

[10] Kowalski (2015) empirically investigates, in the context of 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.

[11] In the quantitative analysis we will introduce a second, fully insured (by assumption) health shock to provide a more accurate map between our model and the health expenditure data.

Assumption 2 *f is continuous, g is twice differentiable and $g'(h) > 0, g''(h) < 0$, $\forall h \in H$.*

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 *If $e' > e$ then $Q(h', h; e)$ first order stochastically dominates $Q(h', h; e')$.*

By assumption, health transitions depend on the effort e of individuals, but not on medical expenditures.[12]

A individual with health status h and current health shock ε that consumes health expenditures x produces $F(h, \varepsilon - x)$ units of output.

Assumption 4 *F is continuously differentiable in both (h, y), increasing in h, and satisfies $F(h, y) = F(h, 0)$ for all $y \leq 0$, and $F_2(h, y) < 0$ as well as $F_2(h, \bar{\varepsilon}) < -1$. Also $F_{22}(h, y) < 0$ 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 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 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 a given 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, 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 on health according to $x = x(\varepsilon, h)$ are spent. Now production and

[12]Our assumption is consistent with the findings by Baiker, et al. (2013). Using the 2008 Medicaid expansion in Oregon they find no significant improvement in objective measures of health (such as blood pressure, cholesterol, and hemoglobin), despite the increased utilization of medical services from Medicaid expansion.

[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).

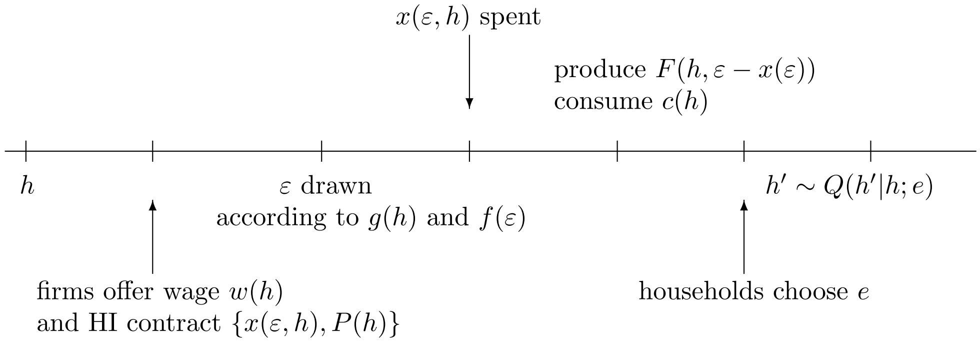


Figure 2: Timing of the Model

consumption takes place. These steps comprise the static, within period-part of the model. 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

There are a large number of production firms that 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, 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 maximize period utility of the household, subject to the firm breaking even.[14] In the absence of government intervention a firm specializing on workers of type h then offers a wage $w^{CE}(h)$ (where CE stands for competitive equilibrium) and 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 health status risk associated with h. This risk stems both from the dependence of wages $w(h)$ as well as health insurance premia $P(h)$ on h, and these are exactly the sources of consumption risk that government policies preventing wage discrimination and prohibiting prior health conditions to affect insurance premia are designed to tackle.

For future reference we also note the obvious result that, since $w(h), x(\varepsilon, h), P(h)$ is chosen to maximize $U^{CE}(h)$ for each health status h separately, the problem (1)-(3) is equivalent to solving

$$\sum_h \Phi(h)U^{CE}(h) = \max_{w(h), x(\varepsilon,h), P(h)} \sum_h u\left(w(h) - P(h)\right)\Phi(h) \tag{4}$$

subject to constraints (2) and (3).

[14] Note that instead of assuming that firms completely specialize by hiring only a specific health type of workers h we could alternatively consider a market structure in which all firms are representative in terms of hiring workers of health types according to the population distribution and pay workers of different health h differential wages according to the schedule $w^{CE}(h)$. In other words 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.

3.4 Government Policies

We now describe how we model policies outlawing health insurance premia to be conditioned on prior health status h, and that limits the extent to which firms can pay workers of varying health h differentially.

3.4.1 No Prior Conditions Law

Under this law health insurance companies face constrains in their pricing, their insurance schedule offers and their applicant acceptance criteria. The purpose of these constraints is to prevent the companies from differentially pricing insurance based upon health status.[15] To be completely 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 assures that the equilibrium health insurance schedule $x(\varepsilon, h)$, given the constraints, is efficient. We now describe the regulations sufficient to achieve this goal.

In order for the no-prior conditions law to be effective the government needs to prevent the emergence of a separating equilibrium in which the health insurance companies (or the production firms in case they offer health insurance contracts) use the health expenditure schedule $x(\varepsilon, h)$ to select their desired health types, given that they are barred from conditioning the health insurance premium P on h directly. Therefore, to achieve pooling in the health insurance market requires the government to regulate the health expenditure schedule $x(\varepsilon, h)$. To give the legislation the best chance of being beneficial we will assume that the government *regulates the health expenditure schedule* $x(\varepsilon, h)$ *efficiently.* For the same reason, since risk pooling is limited if some household types h choose not to 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

$$P = \sum_h \left[g(h)x(0,h) + (1 - g(h)) \int f(\varepsilon)x(\varepsilon, h)d\varepsilon \right] \Phi(h) \tag{5}$$

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. As with the no prior conditions law, the purpose of this legislation is to help insure workers against their 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 effectively circumvent the wage nondiscrimination 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 dire punishment) in achieving its goal of preventing differential hiring and compensation.[16]

Under this legislation then firms will simply take as given the economy-wide wage w^* at which it can hire a representative worker, and the government regulates the insurance market determining the extent of coverage by health type, $x(\varepsilon, h)$, subject to the requirement that the offered health insurance contracts exactly 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).

Perfect competition drives down equilibrium profits of firms to zero with equilibrium wages given by

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

[15] Consistent with this restricted purpose, we also assume that the government cannot use health insurance policies to offset underlying differences in productivity coming from, say, education. This will prove important in the quantitative section.

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

The insurance premium is determined by equation (2), and consumption by $c(h) = w^* - P(h)$ in the absence of a no-prior conditions law. In its presence the premium is given by (5) and consumption, $c(h) = w^* - P$.

Given a cross-sectional health distribution Φ the efficiently regulated insurance contract $x(\varepsilon, h)$ solves:

$$\max_{x(\varepsilon,h)} \sum_h u(w^* - P(h))\Phi(h)$$

subject to (6) and (2) if the no-prior conditions restriction is not imposed on health insurance, and subject to (5) instead of (2) if the no-prior conditions restriction is present. We now turn to the theoretical analysis of the model without and with these policies.

4 Theoretical Analysis of the Model

Given the timing assumptions and the restriction to static health insurance contracts, the analysis of the model can be separated into a static component that, for a given distribution of health Φ, determines wages and the optimal health insurance contract (and where there are no incentives to exercise, $e \equiv 0$), and a dynamic component that determines the optimal health effort choice which leads, through the transition function $Q(h', h; e)$, to stochastic updating of individual health status and thus a new health distribution Φ' at the beginning of the next period. In the next section we analyze the static part of the model before turning to the dynamics in section 4.2.

4.1 Static Analysis

In the analysis of the static component of the model we will characterize competitive equilibrium allocations in the absence and presence of both policies. In order to isolate the sources of inefficiency in the competitive equilibrium (insufficient consumption insurance) that justify government intervention, we first establish the efficient benchmark by analyzing the solution to the social planner problem. The key result in this section is that, statically, the combination of both policies is ideally suited to provide full consumption insurance in the regulated market equilibrium, and thus restores full efficiency of the market outcome.

4.1.1 Social Planner Problem

In order to isolate the potential inefficiencies of the competitive equilibrium allocations in the static model we first analyze an efficient allocation by characterizing the solution to the social planner 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 alternative interpretation, utilitarian social welfare). The planner 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:

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

We summarize the complete optimal solution to the static social planner 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 problem* $\{c^{SP}(\varepsilon,h), x^{SP}(\varepsilon,h)\}_{h\in H}$ *is given by* $x^{SP}(\varepsilon,h) = \max\left[0, \varepsilon - \bar{\varepsilon}^{SP}(h)\right]$, *where the uniquely determined cutoffs* $\{\bar{\varepsilon}^{SP}(h)\}$ *satisfy* $-F_2(h, \bar{\varepsilon}^{SP}(h)) =$

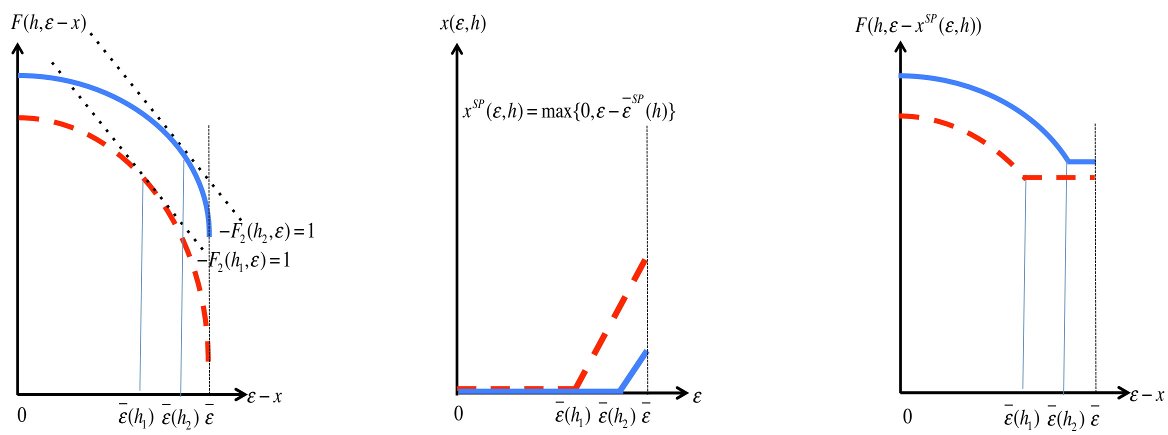
1, *and the first best consumption level is given by*

$$c^{SP}(\varepsilon,h) = c^{SP} = \sum_h \left[g(h)F(h,0) + (1-g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) \left[F(h, \varepsilon - x^{SP}(\varepsilon,h)) - x^{SP}(\varepsilon,h) \right] d\varepsilon \right] \Phi(h). \quad (7)$$

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

Not surprisingly, the social planner finds it optimal to provide full consumption insurance against adverse health shocks ε, but also against bad prior health conditions as consumption c^{SP} is independent of health status 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, given the assumptions on F, to the simple cutoff rule and comparative statics in the proposition.

The efficient level of health expenditure and its implications on production is graphically presented in Figure 3. As shown in the previous proposition, optimal medical expenditures take a simple cutoff rule: small health shocks $\varepsilon < \bar{\varepsilon}^{CE}(h)$ are not treated at all, but all larger shocks are fully treated up to the threshold $\bar{\varepsilon}^{CE}(h)$. 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 $\bar{\varepsilon}^{CE}(h)$ health 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 ($F_{12}(h,y) > 0$, reflected as a "more concave" curve for h_1 than for h_2 in Figure 3(a)), then the equilibrium features better insurance for less healthy households, in the sense of undoing more of the negative health shocks ε through medical treatment $x(\varepsilon,h)$. This is reflected in a lower threshold (more insurance) 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 $F(h, \varepsilon - x^{SP}(\varepsilon,h))$ as shown in Figure 3(c).



(a) Production Function (b) Optimal Medical Expenditures (c) Net-of-Health-Treatment Production

Figure 3: Optimal Medical Expenditures and Production

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

$$U^i(\Phi) = \max_{w(h),x(\varepsilon,h),P(h)} \sum_h u\left(w(h) - P(h)\right) \Phi(h) \quad (8)$$

subject to equations (2) and (5) for premia and equations (3) and (6) 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 condition legislation ($i = NP$), replaces constraint (2) with (5), and the no wage discrimination law ($i = NW$), (3) with (6). 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 (5) and (6) 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 these.

Characterization of the Unregulated Equilibrium ($i = CE$) In section 4.1.1 we saw that the efficient health expenditure allocation is characterized by a simple cutoff rule and consumption is fully insured against health status variations h. 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*

$$-F_2(h, \bar{\varepsilon}^{CE}(h)) = 1 \tag{9}$$

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

Corollary 7 *The equilibrium health expenditure allocation is efficient: $x^{CE}(\varepsilon, h) = x^{SP}(\varepsilon, h)$ for all (ε, h). The cutoff $\bar{\varepsilon}^{CE}(h)$ is increasing in h, strictly so if $F_{12}(h, y) > 0$.*

This corollary shows that in the static case the *only* source of inefficiency of the competitive equilibrium comes from the inefficient lack of consumption insurance against adverse prior health conditions h. This can be seen by noting that

$$\begin{aligned} c^{SP} &= \sum_h \left\{ g(h)F(h,0) + (1-g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) \left[F(h, \varepsilon - x^{SP}(\varepsilon, h)) - x^{SP}(\varepsilon, h) \right] d\varepsilon \right\} \Phi(h) \\ &= \sum_h \left[w^{CE}(h) - P^{CE}(h) \right] \Phi(h) = \sum_h c^{CE}(h)\Phi(h) \end{aligned}$$

and thus aggregate consumption is the same in the equilibrium and efficient allocations.

The equilibrium allocation of health expenditures is efficient, due to the fact that 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 soultion 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. We analyze the behavior of gross wages $w^{CE}(h)$ with respect to health status further in Appendix C, where we provide a sufficient condition for the gross wage schedule to be monotonically increasing in h.

So far we have assumed that the production firms offer a combination of a wage and a health insurance contract to the worker, and have characterized that joint contract. We now briefly show that we can obtain the same allocation through *separate* wage contracts (offered by competitive production firms) and health insurance contracts (offered by competitive health insurers). To do that, 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 or not. 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 8 *A wage contract of the form*

$$\boldsymbol{w}(h,\varepsilon-x)=\begin{cases} w^{CE}(h) & \text{if } F(h,\varepsilon-x)\geq F(h,\bar{\varepsilon}^{CE}(h)) \\ w^{CE}(h)-\left[F(h,\bar{\varepsilon}^{CE}(h))-F(h,\varepsilon-x)\right] & \text{if } F(h,\varepsilon-x)<F(h,\bar{\varepsilon}^{CE}(h)) \end{cases} \tag{10}$$

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.

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.[17]. Also note that an individual that faces this wage contract but cannot buy health insurance will still find it optimal to individually make the competitive health insurance expenditures $x^{CE}(\varepsilon,h)$.

Given the results in this subsection it is plausible to expect, within the context of the static model, that policies preventing competitive equilibrium wages $w^{CE}(h)$ to depend on health status (a wage non-discrimination law) and insurance premia $P^{CE}(h)$ to depend on health status (a no prior conditions law) will restore full efficiency of the policy-regulated competitive equilibrium by providing full consumption insurance. We will 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 Condition Law ($i=NP$) We start with 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 (5). The next proposition characterizes the resulting regulated equilibrium allocation:

Proposition 9 *The equilibrium health expenditures under a no-prior condition 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*

$$-F_2(\tilde{h},\bar{\varepsilon}^{NP}(\tilde{h}))=\frac{\sum_h u'(w^{NP}(h)-P^{NP})\Phi(h)}{u'(w(\tilde{h})-P^{NP})}. \tag{11}$$

The equilibrium wage, for each health status h, is given by equation (3), and the health insurance premium is determined by equation (5), 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 10 *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 $-F_2(\tilde{h},\bar{\varepsilon}^{NP}(\tilde{h}))=1$.*[18] *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.*

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

[18] For the purpose of the proposition it does not matter whether $\tilde{h}\in H$ or not.

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 in the labor market through a no discrimination law. The allocative consequences of the law are summarized in the following proposition whose proof follows directly from the first order conditions of the program (8).

Proposition 11 *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*

$$-F_2(\tilde{h}, \bar{\varepsilon}^{NW}(\tilde{h})) = \frac{u'(w^{NW} - P^{NW}(\tilde{h}))}{\sum_h u'(w^{NW} - P^{NW}(h))\Phi(h)}. \tag{12}$$

The equilibrium wage is determined by equation (6) and the health insurance premium is given by, for each h, by 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 $\bar{\varepsilon}^{NW}(h)$. Note that 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 forces, preventing us from establishing monotonicity in cutoffs $\bar{\varepsilon}^{NW}(h)$ across health groups h. On one hand, a one unit increase in medical expenditure $P(h)$ is more costly to the unhealthy since 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.[19]

Competitive Equilibrium with Both Policies ($i = B$) Finally, combining both a no-wage discrimination law and a no-prior conditions legislation restores efficiency of the regulated equilibrium since both policies in conjunction provide full consumption insurance against bad health realizations h. This is the content of the next proposition.

Corollary 12 *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.*

The no prior conditions law equalizes health insurance premia P across health types, the no wage discrimination law implements a common wage w^* across health types, and the (assumed) efficient regulation of the health insurance market assures that the health expenditure schedule is efficient as well.

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. Both a no-prior conditions law and a no-wage discrimination law provide partial, but not complete, consumption insurance against this risk, without distorting the effort level. 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 first best allocation in the static model. Enacting both policies jointly is thus fully successful in what they are designed to achieve in a static world (partially due to the fact that additional government regulation severely restricted the options of firms to circumvent the government policies).

[19] The optimal cutoff condition does imply that agent's with high marginal utility relative to the average will have lower cutoffs, and vice versa. So long as these agents still correspond to the less and more healthy agents, then we can say that the cutoffs for the less healthy have been distorted downward for the more healthy agents and upward for the less health relative to the social optimum.

4.2 Analysis of the Dynamic Model

We now study the full dynamic version of our economy. Both in terms of casting the problem, as well as in terms of its computation we make use of the fact that there is no aggregate risk. Therefore the sequence of cross-sectional health distributions $\{\Phi_t\}_{t=0}^T$ is a deterministic sequence. Furthermore, conditional on a distribution 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:

$$\Phi_{t+1}(h') = \sum_h Q(h'; h, e_t(h))\Phi_t(h). \tag{13}$$

Under each policy, given a sequence of aggregate distributions $\{\Phi_t\}_{t=0}^T$ we can solve an appropriate dynamic maximization problem of an individual household for the sequence of optimal effort decisions $\{e_t(h)_{h\in H}\}_{t=0}^T$. For this, in this section we assume that 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 of the model.

A sequence of optimal effort choices in turn implies a new sequence of aggregate distributions via (13). 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 Problem

As in the static model, we first study, as point for comparison for equilibrium allocations (without and with policies), the solution of a planner problem determining constrained-efficient allocations. In the static model the planner can provide full consumption insurance against initial health conditions, as could both policies. In the dynamic model with endogenous effort choice a fully 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 we believe it is more instructive for comparison to study a restricted dynamic planner problem in which the social planner has to respect the intertemporal optimality condition with respect to private 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 directly 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.

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

$$\max_{\{c_t(h), e_t(h)\}} \quad V(\Phi_0) = \sum_h V_0(h)\Phi_0(h) \tag{14}$$

$$s.t. \quad \sum_h c_t(h)\Phi_t(h) \tag{15}$$

$$\leq \sum_h \left[g(h)F(h,0) + (1-g(h)) \int_\varepsilon f(\varepsilon)\left[F(h, \varepsilon - x^{SP}(\varepsilon,h)) - x^{SP}(\varepsilon,h)\right] d\varepsilon \right] \Phi_t(h)$$

$$q'(e_t(h)) = \beta \sum_{h'} \frac{\partial Q(h'; h, e_t(h))}{\partial e_t(h)} V_{t+1}(h'), \tag{16}$$

and the general law of motion given in (13). Equation (15) represents the aggregate resource constraint, where $x^{SP}(\varepsilon, h)$ is the efficient health expenditure schedule[20] characterized in section 4.1.1. The second

[20]Since health expenditures only affect current productivity and do not interact with current effort choice, there is no gain for the planner to deviate from static production efficiency.

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 onward for that household, *given* a consumption allocation chosen by the planner and encoded in $\{V_{t+1}(h')\}$.

Equation (16) demonstrates the basic trade-off for the constrained dynamic social planner that will also be present in our evaluation of both policies. In the static analysis we showed that is was optimal for the social planner to provide full consumption insurance. Although the constrained planner can certainly implement such an allocation, it will lead to identically zero effort in all periods, on account of equation (16).

Proposition 13 *Suppose the planner chooses a consumption allocation $\{c_t(h)\}$ satisfying $c_t(h) = c_t$ for all $h \in H$. Then effort is identically equal to zero, $e_t(h) = 0$ for all t, h.*

Since the marginal cost of providing effort at $e_t(h) = 0$ is zero (assumption 1) and the benefit on health transitions and thus net production and average consumption is positive (on account of assumptions 3 and 4), starting from the full consumption insurance, 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), and thus $e_t^{SP}(h) > 0$. This result will be in contrast to the outcome under both policies (again see proposition 14) below which[21] features full consumption insurance and zero effort, and therefore results in an inefficient allocation, relative to the (constrained) social planner solution.

4.2.2 Dynamic Competitive Equilibrium without and with Policy

In our model, absent wage and health insurance policies households do not interact in any way, and in the presence of such policies they interact since the cross-sectional health distribution Φ_t determines the pooled wage and/or health insurance premium, and the optimal health expenditure cutoff $\bar{\varepsilon}^i(h; \Phi_t)$ in policy regime i. The dynamic program in policy regime $i \in \{CE, NW, NP, B\}$ reads as:

$$v_t^i(h; \Phi) = U^i(h, \Phi) + \max_{e_t(h)} \left\{ -q(e_t(h)) + \beta \sum_{h'} Q(h'; h, e_t(h)) v_{t+1}^i(h', \Phi') \right\} \tag{17}$$

where

$$U^i(h, \Phi) = \max_{x^i(\varepsilon,h,\Phi), w^i(h,\Phi), P^i(h,\Phi)} u(w^i(h, \Phi) - P^i(h, \Phi)) \tag{18}$$

and

$$P^i(h; \Phi) = (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) x^i(\varepsilon, h; \Phi) d\varepsilon \text{ if } i \in \{CE, NW\} \tag{19}$$

$$P^i(\Phi) = \sum_h \left[(1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) x^i(\varepsilon, h; \Phi) d\varepsilon \right] \Phi(h) \text{ if } i \in \{NP, B\} \tag{20}$$

$$w^i(h; \Phi) = g(h) F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) F(h, \varepsilon - x^i(\varepsilon, h; \Phi)) d\varepsilon \text{ if } \{CE, NP\} \tag{21}$$

$$w^i(\Phi) = \sum_h \left\{ g(h) F(h, 0) + (1 - g(h)) \int_0^{\bar{\varepsilon}} f(\varepsilon) [F(h, \varepsilon - x^i(\varepsilon, h; \Phi))] d\varepsilon \right\} \Phi(h) \text{ if } i \in \{NW, B\} \tag{22}$$

and $x^i(\varepsilon, h; \Phi)$ is the equilibrium health expenditure allocation solution to the static equilibrium problem in section 4.1.2 if policy regime i is in place.

Recall from section 4.1.2 that the optimal and equilibrium health expenditure allocation was given by the simple cutoff rule

$$x^i(\varepsilon, h; \Phi) = \max \left[0, \varepsilon - \bar{\varepsilon}^i(h, \Phi) \right] \tag{23}$$

[21] This statement is true absent direct utility benefits from better health, as assumed in the theoretical model but relaxed in the quantitative implementation.

and that the policy dependent cutoff rule satisfies, for all $\tilde{h} \in H$ and for the policy regimes $i \in \{CE, NP, NW, B\}$

$$-F_2(\tilde{h}, \bar{\varepsilon}^{CE}(\tilde{h})) = -F_2(\tilde{h}, \bar{\varepsilon}^{B}(\tilde{h})) = 1 \tag{24}$$

$$-F_2(\tilde{h}, \bar{\varepsilon}^{NP}(\tilde{h}; \Phi_t)) = \frac{\sum_h u'(w^{NP}(h;\Phi_t) - P^{NP}(\Phi_t))\Phi_t(h)}{u'(w^{NP}(\tilde{h};\Phi_t) - P^{NP}(\Phi_t))} \tag{25}$$

$$-F_2(h, \bar{\varepsilon}^{NW}(\tilde{h}; \Phi_t)) = \frac{u'(w^{NW}(\Phi_t) - P^{NW}(\tilde{h}, \Phi_t))}{\sum_h u'(w^{NW}(\Phi_t) - P^{NW}(h, \Phi_t))\Phi_t(h)} \tag{26}$$

Thus, whether, and how, $x^i(\varepsilon, h; \Phi)$ and $w^i(h;\Phi), P^i(h;\Phi)$ and thus consumption $c^i(h,\Phi) = w^i(h,\Phi) - P^i(h,\Phi)$ depend on the cross-sectional health distribution Φ in turn varies across policy regimes i, as we saw in section 4.1.2. However, note that in order for the household to solve her dynamic programming problem she 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

$$v_t^i(h) = u(w_t^i(h) - P_t^i(h)) + \max_{e_t(h)} \left\{ -q(e_t(h)) + \beta \sum_{h'} Q(h'; h, e_t(h)) v_{t+1}^i(h') \right\} \tag{27}$$

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

$$q'(e_t(h)) = \beta \sum_{h'} \frac{\partial Q(h'; h, e_t(h))}{\partial e_t(h)} v_{t+1}^i(h'). \tag{28}$$

and thus equates the marginal cost of providing effort, $q'(e)$ with the marginal benefit of an improved health distribution tomorrow. By assumptions 1 and 3, from equation (28) 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'.

Although equation (28) looks identical across policy regimes, the determination of the value functions that appear on the right hand side of both equations is not. Comparing equations (20) to equation (19), and (22) to equation (21) highlights the extra consumption insurance induced by the no-prior conditions and wage nondiscrimination laws. This extra consumption insurance, 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.

We make the following observations. First, in the unregulated equilibrium, $i = CE$, since the equilibrium expenditure cutoff and thus the health spending allocation of a health type h is independent of the cross-sectional health distribution, there is no interaction of health types at all.

Second, with one or both policies in place, 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 shown in equations (20) and (22), as well as through the health expenditure cutoffs, see equations (25) and (26). Both through wage and premium pooling as well as distorting health expenditure allocations, the policies provide partial consumption insurance against health risk. In the dynamic model with endogenous effort choice, this happens at the expense of providing 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 14 *Suppose there is a no wage discrimination and a no prior condition 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* (13) *with $e_t(h) \equiv 0$.*

Corollary 15 *The allocation under both policies is inefficient since health effort is inefficiently low.*

In the presence of both policies there are no incentives, either through wages or health insurance premia, to exert effort to lead a healthy life. Since effort is costly, households will not provide any such effort in

[22] In appendix E we describe our computational algorithm to solve the dynamic competitive equilibrium model that exploits this insight.

the regulated dynamic competitive equilibrium. Thus in the absence of any direct utility benefits of better health the combination of both policies leads to a complete collapse in incentives, with the associated adverse long run consequences for the distribution of health in society.

Equipped with these theoretical results and the numerical algorithms to solve our model we now map our model to cross-sectional health and exercise data from the PSID to quantify the effects of government regulations on the evolution of the cross-sectional health distribution, as well as aggregate production, consumption and health expenditures.

5 Mapping the Model to the Data

Prior to spelling out the details of the empirical implementation it is useful to review the basic logic of our model, in order to highlight what key aspects of the data will drive our parameterization of the model and thus in turn the quantitative results. It will also serve to justify the extensions of the model for the empirical analysis discussed in the next subsection.

First, as displayed in section 2.1, in the data we observe a *positive correlation between health and wages*. In the model the size of the impact of health status on output and thus wages is determined by the production function for final output, $F(h, \varepsilon - x)$. How strongly wages depend on health will determine both the scope of the benefits of insuring wages against health-induced fluctuations as well as the efficiency costs of a deteriorating population health distribution.

Second, the data in section 2.1 display a *negative correlation between health status and health expenditures* which will inform the estimation of the probability $g(h)$ of receiving a productivity-reducing health shock ε. The strength of this dependence in turn will quantify the insurance benefits and efficiency cost of a no prior conditions legislation restricting the dependence of health insurance premia on health status h.

The *third* key observation from section 2.1 is that, controlling for current health, empirical *measures of health effort e raise, in expectation, future health status*. In the model the magnitude of this effect is encoded in the health transition function Q. Therefore the three "technology" functions $F(h, \varepsilon - x), g(h), Q(h'|, h, e)$ capture the impact of health status and health efforts on economic outcomes, and the utility and cost functions $u(c), q(e)$ measure its welfare consequences.

5.1 Extending the Model

To obtain an empirically plausible magnitude of the three key effects discussed above in the model, and provide a satisfactory representation of the micro data we enrich the model in four dimensions.

First, in the data, the strong association of wages and health spending with health status can be driven either by the direct impact of health, or by other observable or unobservable factors correlated both with health status and income as well as health expenditures. It is crucial for our analysis to get the direct impact of health on these two variables right. Therefore, 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 F.4 and F.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 our quantitative 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 with completed high school or less, and individuals with at least some college education. Conditioning on observables only addresses the concern that individuals across different health status bins $h_1, .., h_4$ possess different characteristics independently affecting wages and health expenditures. In section 7.1.1 we investigate the issue of unobserved (to the econometrician) heterogeneity across health status cells, 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) by arguing that impacts of health on wages implied by our model calibration estimation are consistent with the empirical estimates from the literature.

Second, in the model analyzed so far the only purpose of medical expenditures is to offset labor productivity reducing health shocks ε. In the data, in contrast, some households have health expenditures in a given year from catastrophic illnesses that exceed their labor earnings. These exceedingly rare but large expenditures cannot be rationalized by the benchmark model. In order to arrive at realistic magnitudes of health insurance premia in the augmented model 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. 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 $\mu_z(t, educ, h)$.

Third, in the model health effort e is a one dimensional variable chosen by households to balance the utility cost of effort and the benefits of better health (in expectation) in the future. In the PSID data we observe three measures of health efforts (light exercise, heavy exercise, the degree of abstaining from smoking). To accommodate the empirical fact that effort is multi-dimensional in the data we now assume that overall effort e is a weighted average of the different types of health efforts. In section 7.1.3 we further extend the model to allow for other unobserved (and thus mis-measured) components of effort and demonstrate that our policy conclusions are robust to health effort being observed with this measurement error in the data.

Fourth, in the model thus far all households have the same disutility of this health effort. The data suggests that households with worse health status tend to exercise less; in order to help capture this empirical regularity we now augment the model with a health-specific preference shifter $\gamma(h)$ so that the cost of health 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 (which determines the productive benefits of better health h) remains completely unchanged (and so do the optimal health insurance contracts and health expenditure allocations). In the dynamic analysis only the marginal cost of effort changes, to $\gamma(h)q'(.)$ but the analysis is otherwise unaltered.

Finally, the model thus far assumed that the continuation utility at retirement was independent of the health status h at retirement, and thus, 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 will govern the incentives to provide health effort late in working life (which otherwise would be zero) and will be within the model to match the health effort profile of older workers. This extension of the model also avoids the complete collapse of incentives under both policies (that is, proposition 14 from the theoretical model no longer applies). It captures, albeit in a fairly reduced-form way, 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, we attempt, through the terminal value function $v_{T+1}(educ, h)$ to include the alternative health utility benefits part of the literature has modeled explicitly (see e.g Hall and Jones (2007) or Ales et al., 2012). We now turn to the determination of the parameters of the augmented model that we will use for the policy experiments.

5.2 Parameter Estimation

We assume that one model period is six years, a compromise between assuring 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).[23] We then divide the set of model parameters into three broad categories, a small subset of preference parameters that we choose outside the model, a second set of health expenditure and health transition parameters governing $Q(h'|h, e; educ), \mu_z(t, educ, h)$ 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* $\beta = 0.96$.

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

5.2.2 Parameters Estimated Directly from the Data

In the second step we estimate part of the model parameters directly from the data, without having to rely on information from the equilibrium of the model. First, we can deduce the initial cross-sectional health-education distribution $\Phi_0(h, educ)$ directly from the PSID data. We obtain the following additional model elements also directly from the data.

Health Transition Function $Q(h'|h, e; educ)$ With our assumptions the only endogenous choice affecting health status decisions is the effort choice e. Thus with panel data on health status that also contains information on empirical proxies for costly health efforts the Markov transition function Q can be estimated directly from the data. The PSID, starting from 1999, contains measures of light exercise e^l, heavy exercise e^h as well as the number of cigarettes smoked in a day s. We normalize[24] each measure to a range between 0 and 1 and then estimate the following parametric form for the health transition function, with all parameters being permitted to be education-specific (a dependence that is suppressed in the notation below):

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

where $e = (\delta_l e^l + \delta_h e^h + \delta_s(1-s)), \quad \delta_l + \delta_h + \delta_s = 1$.

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 transition probability function 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 between today and tomorrow. Since smoking has adverse effects on health, we use $1 - s$ as a measure of *not* smoking effort. Moreover, light, heavy exercise and smoking may have different effects on health transitions, and thus we permit the weights $\{\delta_l, \delta_h, \delta_s\}$ for each effort component to differ. We think of $e = \delta_l e^l + \delta_h e^h + \delta_s(1-s)$ as the effort variable used in the theoretical analysis of our model.

At the core of our paper are the incentive effects on individual effort of social insurance interventions. The cross-sectional heterogeneity in effort-induced health transitions provides crucial evidence along this dimension, and thus we aim to fully capture this heterogeneity. We therefore estimate this process using a (completely standard) Maximum Likelihood procedure on individual data described in detail in Appendix F.3. Figure 4 and table 4 provide a summary of the estimated health transition functions, and how they depend on the observed health effort choices. These are one of the three quantitative key model ingredients, but might be of independent interest. In figure 4, each initial health status $h \in \{h_1, .., h_4\}$ accounts for one of the four sub-figures, and it plots the estimated probabilities as well as data-implied frequencies[25] of transitioning into health status $h' \in \{h_1, .., h_4\}$ as function of health effort e. The estimated transitions capture well that, qualitatively, in the data higher health effort is associated with a larger frequency of favorable health transitions, and for most (h, h') pairs, captures the gradient with respect to e well. The

[24] Light exercise is measured as the number of days an individual carries out light physical activity (walking, dancing, gardening, golfing, bowling, etc.); 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 activity is performed. For smoking, we use the number of cigarettes smoked per day with 50 cigarettes being the maximum and normalized to 1.

[25] The lines are the model-implied, estimated transition probabilities. In order to summarize the underlying raw data on which these estimates are based, in the data, for each initial h we group all households with that h into five groups according to their observed health effort choices and scatter-plot their transition frequencies. The size of each (diamond, circle, star, square) represents the number of observations in the respective group. Figure 4 contains weighted averages of transition probabilities across education groups. The estimated transition functions by education group are contained in appendix F.3.

exceptions are concentrated among e groups with few observations that the maximum likelihood procedure -which uses the 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 $\}$

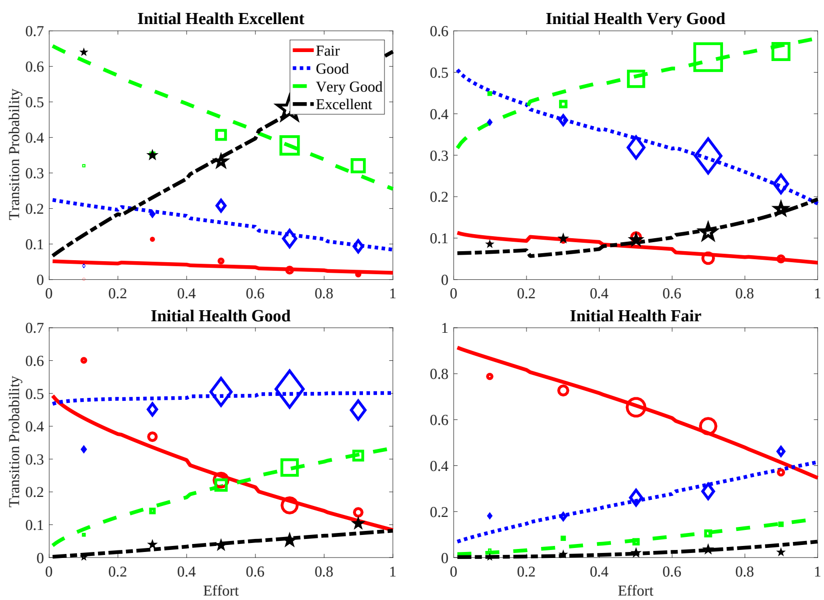


Figure 4: Health Transition Functions - Data (Scatter) vs. Estimated (Line)

As further justification that the estimated impact of health effort on health transitions is plausible we relate the implications of our estimates to an empirical study that has estimated the impact of health efforts on the prevalence of individual health conditions. Colman and Dave (2013) use the first National Health and Nutritian Examination Survey (NHANES I) conducted between 1971 and 1974 and its follow-up study, the NHANES I Epidemiologic Follow-up Study (NHEFS) spanning the years 1982 to 1984 to estimate the impact of physical activity on the risk factors for heart disease. Using exercise survey (reported as one of very active, moderately active, or quite inactive), they document that relative to low levels of recreational exercise, high levels of exercise reduce the probability of hypertension, heart disease, and diabetes by 8.4, 2.4, and 2.2 percentage points, respectively.[26]

Using data from MEPS, summarized in table 20 in Appendix F.3 we can map our measure of health status h into the probability of a specific health condition (e.g. hypertension or diabetes). We then divide our PSID sample into four health status and three effort groups and use our estimated health transition

[26]Moderately active individuals experienced about 5.1, 1.9, and 1.8 percentage point decline in the risk of hypertension, heart disease, and diabetes.

function to determine the impact of health effort e on the incidence of a specific health condition in the future implied by our estimated Q (through the health status transitions and the map between health status and medical conditions). Using this procedure we find that our estimated health transition function predicts lower incidence of a specific health condition of about 2-3 percentage points among the high- relative to the low exercise group, consistent but at the low end of the direct empirical estimates of Colman and Dave (2013). This, in our view, provides further evidence that the quantitative impact of health efforts of health transitions is empirically plausible.

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

5.2.3 Parameters Estimated within the Model

We estimate the remaining parameters of the model via the generalized method of moments (GMM) so that selected model-equilibrium implied statistics match their empirical counterpart.[27] The parameters that remain to be estimated include those governing the production function $F(t, educ, h, \varepsilon - x)$, the health shock process (the probability of receiving a shock $g(t, h)$ and the shock distribution $f(t, h)$) and the remaining preference parameters (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.

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]
$\mu_\varepsilon, \sigma^2_\varepsilon$	[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

[27] 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 health-status related elements of our model into the data we opt for an empirical GMM approach that matches health-group averages in the data, rather than attempting to exploit the pervasive individual variation in wages though maximum likelihood, as we did for the health transition function.

Note that the structure of our model allows us to estimate the parameters in two separate steps. Since realized wages and health expenditures are determined exclusively from the *static* part of the model and are independent of effort decisions and the associated health evolution in the dynamic part, in a first step we separately 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 preference parameters.

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

We assume that the production technology takes the following functional form:

$$F(t, educ, h, \varepsilon - x) = A(t, educ, \tilde{h}) h^{\alpha_t(t)\alpha_e(educ)} - A(t, educ, \tilde{h}) \phi(educ) \frac{(\varepsilon - x)^{\xi(educ)}}{h} \tag{29}$$

with $\tilde{h} \in \{\{h_1\}, \{h_2, h_3, h_4\}\}$. This specification of the production function encodes two impacts of health status on productivity and thus labor income. The first term in equation equation (29) captures the impact of health status (and age and education) on wages in the *absence of health shocks* and thus health expenditures ($\varepsilon = x = 0$). The base level of wages $A(t, educ, \tilde{h})$ is determined by fully flexible age and education effects, and permits a wage discount for the lowest health group (as 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, standard, component of wages in the model is to capture, in a flexible way, the mean life cycle profiles of wages and their dependence on education and health *status* in the data.[28] Note that the health status levels $h \in \{h_1, h_2, h_3, h_4\}$ themselves are also parameters to be estimated.[29]

The second part of equation equation (29) 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 $\phi, \xi(educ)$. By dividing this term by the level of health h, 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 magnitude of the parameters $\phi, \xi(educ)$ in part determines how strongly health expenditures x and thus insurance premia increase with lower health status h.

The other model component that determines the health status, health expenditure gradient in the model 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 $\tilde{g}(h)$ summarizes the positive impact of health status, and α_g represents the negative age effect on the this probability. Conditional on receiving an ε shock, the distribution of these shocks 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 $\bar{\varepsilon}(t, educ, h)(h)$. 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 in conjunction determine observed 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

[28] 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.

[29] The categories {Excellent, Very Good, Good, Fair} used in the data itself have no cardinal interpretation.

[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 a log normal distribution to fit their data well. In principle, we could estimate the age-dependent (ε distributions nonparametrically 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 we use here.

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

status on labor productivity and thus incomes as well as on health expenditures is crucial for our quantitative results. The key requirement of the functions $F(\cdot), g(\cdot), f(\cdot)$ is that they are flexible enough to match the data well, whereas their exact functional form is secondary. The crucial determinant of the health impact on productivity and expenditures are then the empirical moments we use as targets. We model several key dimensions of household heterogeneity explicitly that may affect both the health distribution as well as wages and health expenditures, notably age and education. We also control for other *observable* factors that may be correlated with health status and potentially affect labor income and medical expenditures (such as race and gender). 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 age group-education-and health dummies to construct the empirical education- and health-specific age income moments. We smooth these empirical moments by fitting them to education- and health-specific quadratic functions of age groups, and use the predicted income and medical expenditure profiles as our moments. Details of this procedure are contained in Appendices F.4 (for wages) and F.5 (for health expenditures).

As expected, conditioning on observables before deriving the health gradient of wages and health expenditures lowers the dependence on health status relative to the raw data, but the dependence remains quantitatively very important, as Figure 5 demonstrates.[32] It displays the empirical targets for wages (equivalently, labor income) as well as medical expenditures over the life cycle, and also includes the model-implied profiles (Table 18 in the Appendix contains the numbers underlying these plots). These match the empirical targets very well, fully acknowledging that this is of course 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.

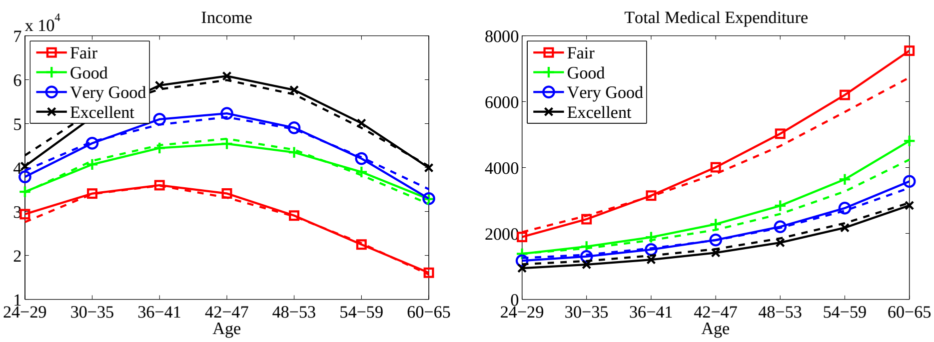


Figure 5: Income and Medical Expenditure - Data (Solid) vs. Model (Dashed)

A final concern about the estimated income and health expenditure profiles could be that they are partially driven by an unobserved factor that in the population is correlated both with health status and income (and/or the incidence of health shocks or effort). To investigate this possibility further we now explore the properties of the regression residuals (after controlling for all observables, including health status) for income, health expenditures and effort. If an underlying unobserved factor drives these variables at the individual level we would expect the regression residuals to be strongly correlated in the cross-section. Instead, we find that these correlations are quite small, suggesting that it is unlikely that an unobservable fixed factor jointly drives these three variables.[33]

One caveat to this discussion is that a high individual fixed factor that only impacts labor productivity increases the incentive of a household to improve her health, leading to higher average productivity among higher h types. This could potentially account for part of the observed income-health gradient even once observables are controlled for. In order to assess the potential concern that this specific type of unobserved

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

[33] For example, 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 is 0.02. Also note that the correlation between income as well as health expenditure residuals and future health is small and negative, strongly suggesting that the causal connection is not running in the reverse direction, from an individual income and expenditures to future health.

heterogeneity is responsible for the results of our policy analysis, in section 7.1.1, as part of our robustness analysis, we shrink the health-status related income gaps (which are already controlled for by observable differences across h groups) by approximately 60% and demonstrate that the ranking of policies is unaffected by this change (although, not surprisingly, the welfare gains from the policies become smaller).

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 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 holds true in our estimation results). The parameter ψ controls the elasticity of the utility cost with respect to effort, $q'(e)/(q(e)/e)$. When ψ increases, the elasticity of the cost of effort increases for all effort levels.[34] The health-dependent effort cost shifters $\gamma(h)$ allow us to account better for the *level* of exercise observed in the data.

As discussed above, we introduce a terminal, education and health dependent continuation utility $v_{T+1}(educ, h)$ to rationalize that even older workers exert health effort in the model. Note that only the difference in the continuation values matter for the choice of optimal effort in the last period T and thus we normalize $v_{T+1}(educ, h_1) = 0$. We estimate these 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[35] 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 17 in Appendix F.6. As with wages and health expenditures our model is parameterized

[34] Empirically, the wage-health 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 wage-health gradients allows us to estimate ψ.

[35] It would be optimal to use 7 age groups, but partitioning the data so finely results is very small cell sizes. Young workers are those aged 41 or less.

flexibly enough to capture the variation of health effort across age, education and health groups well. 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.

6 Results of the Policy Experiments

We now use the quantitative model to answer the main policy question of this paper: what are the effects of introducing the wage nondiscrimination and no-prior condition legislation on the evolution of the distribution of health, consumption and effort, and ultimately, on social welfare? These policies trade off 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 of policies, and second, the adverse incentive effects on aggregate production and health. Then, in subsection 6.3, we display the welfare consequences of our policy reforms.[36] We conclude with various sensitivity analyses in section 7.

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. We observe that the combination of both policies, as predicted by the theory (see section 4.1), is fully effective in providing perfect consumption insurance. As in the first period of the constrained efficient allocation, *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.[37] Also notice from figure 6 that a wage non-discrimination law alone goes a long way towards providing effective consumption insurance, since the effect of health status on labor income is quantitatively larger than 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, the comparison to the constrained efficient allocation suggests that a perfect wage pooling policy provides somewhat too much insurance, whereas the health premium pooling policy alone delivers significantly too little.

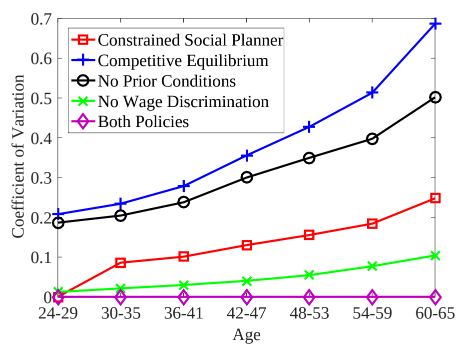


Figure 6: Consumption Dispersion

The consumption insurance described in the previous figure is achieved by 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 theoretically in Proposition 10, the healthy workers subsidize the premium of the unhealthy. Similarly, wages of the unhealthy workers are subsidized by the healthy, productive

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

[37] Due to the presence of heterogeneity in education levels the economy as a whole displays non-trivial consumption dispersion even in the presence of both policies.

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 measures the differences between the actuarially fair health insurance premium a particular health type household 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 productivity of the worker and the wage received under a no-wage discrimination policy and in the presence of both policies. Negative numbers imply that the worker is paying a higher premium, or is paid lower wage than in the unregulated competitive equilibrium. Thus such a worker, in the presence of the policies, has to transfer resources to workers of different (lower) health types. Reversely, positive numbers imply that a worker is being subsidized in her labor income or health insurance premium. To interpret the numbers, the units on the y-axis are in percent of consumption for the specific age-health group in question.

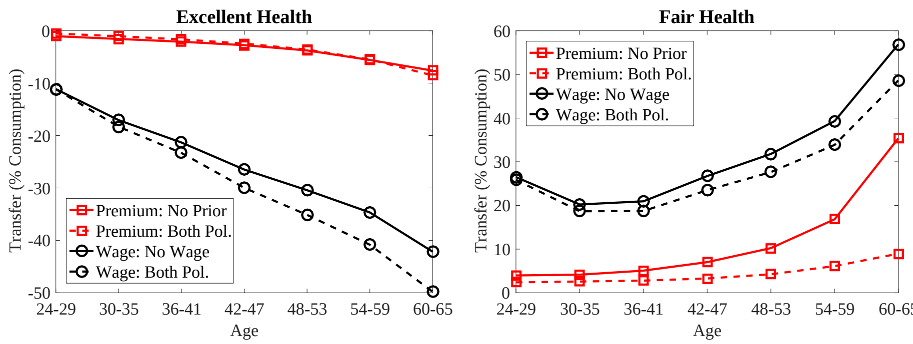


Figure 7: Cross Subsidy: Excellent and Fair Health

From Figure 7 we observe that the workers with excellent health significantly cross-subsidize the other workers, both in terms of health insurance premia as well as in terms of wage transfers. Thus, under the wage non-discrimination legislation young workers with excellent health are paid roughly 10% less than what their productivity would demand, this number rising 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 and thus low labor productivity. In relation to their consumption, the right panel of figure 7 shows that these transfers are quantitatively very large, constituting the main source of consumption especially at older ages (where not only the productivity differentials by between health groups are large, but so is the share of the population with fair health). Figure 7 also demonstrates what is already implicit in figure 6: transfers induced by a no-prior conditions law are significantly smaller than implied by the wage law (not exceeding 10% of consumption for most individuals except for the oldest individuals of fair health), but still a very importance source of insurance, especially for older individuals of poor health.[38]

6.2 Adverse Incentive Effects on Aggregate Production and Health

In this subsection, we now analyze the aggregate dynamic effects of the policies on production and the health distribution. The associated 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

[38] An interesting property of the subsidies is that the level of subsidization implied by a given policy depends somewhat on whether the other policy is present or not. 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 theoretically in section 4.1 with only a no-prior condition law in effect, the resulting health insurance contract is more generous for the unhealthy, in order to provide income insurance through the back door, 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 much larger (especially for old individuals) if only a NP-law is present (35% of consumption vs. 10% of consumption for the oldest age group), as the right panel of figure 7 demonstrates.

policies,[39] but substantially lower in the presence of the non-discrimination laws. 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 induce workers to provide effort in the first place, and thus the policies that reduce that consumption dispersion the most come with the sharpest reduction in incentives.[40] 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 post-retirement continuation utility, and thus effort plummets strongly, relative to the competitive equilibrium.[41]

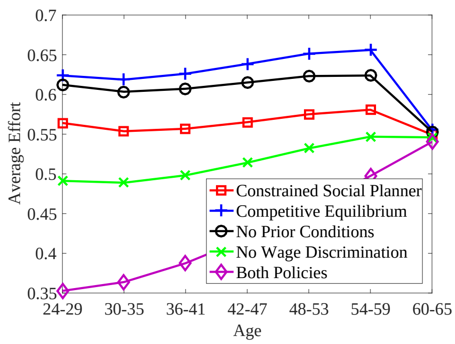


Figure 8: Effort

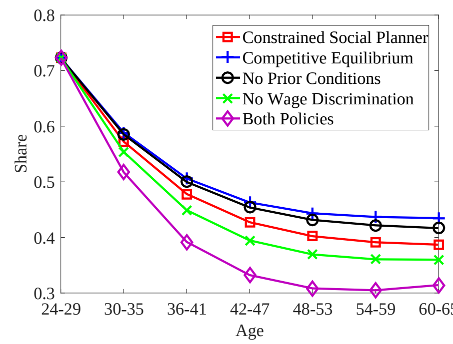


Figure 9: Pop. Share with Excellent, Very Good Health

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 which displays the share of households in the population with excellent and very good health (as one summary measure of the population health distribution), is then 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 nondiscrimination law is in place. As with effort, 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.

This deterioration of the population health distribution has significant impact on macroeconomic aggregates, such as total spending on health, aggregate output and consumption, as figures 10, 12 and 13 demonstrate. For a better comparison across policy regimes, 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. However, the speed of this increase over the life cycle differs significantly across policy regimes.

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[42] by health status h in Figure 11. Recall from section 4.1 that the thresholds $\bar{\varepsilon}(h)$ under the unregulated competitive equilibrium and in the presence of both policies are socially efficient and thus the three graphs completely overlap. 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

[39] Because we have introduced a terminal value of health which induces not only effort in the last period even under both policies, but through the continuation values in the dynamic programming problem, positive effort in all periods.

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

[41] The impact of the policies on effort is most significant at young and middle ages, whereas towards retirement effort levels under all polices converge. This is owed to the fact that 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 through the entire working life and are strongly affected by the policies.

[42] For the youngest age group with low education, they look qualitatively similar for other age-education groups.

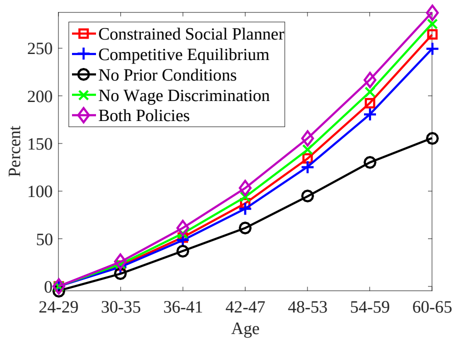


Figure 10: Health Spending

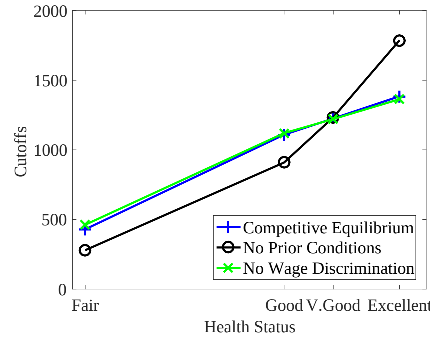


Figure 11: Health Expenditure Cutoffs

under-insured. This is the content of Proposition 10, and it is quantitatively responsible (together with relatively high effort and thus a more slowly declining fraction of individuals with excellent health in the population, see again Figure 9) 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 (relative to the competitive equilibrium) are minor. The larger health expenditures especially later in life in Figure 10 stem primarily from a faster decline of 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 on the non-treated fraction of her health shock, aggregate output is lower, ceteris paribus, under policy configurations that lead to a worse health distribution and that leave a larger share of health shocks ε untreated. From figure 12 we observe that the deterioration of health under a policy environment that includes a wage nondiscrimination 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 condition legislation, entails a significant loss of average consumption in society. Note that, in terms of aggregate consumption, the fact that health expenditures are significantly lower under a no-prior conditions law than in the other policy configurations implies that a larger share is available for consumption. As a consequence, aggregate consumption is actually larger than in the competitive equilibrium early in the life cycle, and not significantly lower in old age.

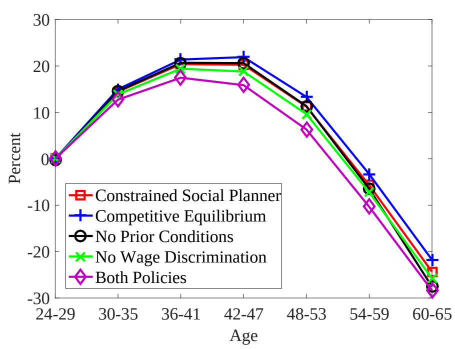


Figure 12: Production

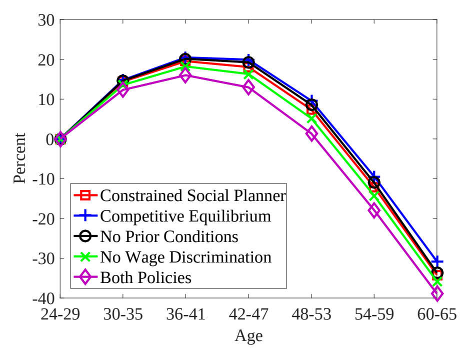


Figure 13: Consumption

Overall, the effect on aggregate effort, health, production and thus consumption suggests a quantitatively important trade-off between consumption insurance and incentives. The unregulated competitive equilibrium provides strong incentives at the expense of risky consumption, whereas a policy mix that includes both policies provides full insurance at the expense of a significant deterioration of the population health distribution. The effects of the no-prior conditions law on both consumption insurance and incentives are modest, relative to the unregulated equilibrium. In contrast, implementing a no wage discrimination law or both policies insures away most of the consumption risk, but significantly reduces (although does not eliminate completely) the incentives to exert effort to lead a healthy life, especially early in the life cycle. The comparison to the constrained efficient allocation suggests that the wage policy resolves this trade-off more successfully, but by *fully* insuring the very significant income risks leads a reduction in health incentives that is too large. We now confirm this by deducing, in the next section, the welfare consequences from hypothetical reforms towards these pure policies. In section 7.3.1 we further strengthen this conclusion by showing that a policy of only partial (but very substantial) wage compression does even better still.

6.3 Welfare Implications

6.3.1 Aggregate Welfare

In this section we quantify the welfare impact of the policy innovations studied in this paper. For a fixed initial distribution $\Phi_0(h)$ over health status,[43] denote by $W(c,e)$ the expected lifetime utility of a cohort member (where expectations are taken prior to the initial draw h of health) from an arbitrary allocation of consumption *and effort* over the life cycle.[44] Our consumption-equivalent measure of the welfare consequences of a policy reform is given by

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

where $i \in \{SP, NP, NW, B\}$ denotes the different policy scenarios. Thus CEV^i is the percentage reduction of consumption in the competitive equilibrium *consumption* allocation required to make households indifferent (ex ante) between the competitive equilibrium allocation[45] 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, again taking as given the initial distribution Φ_0. In the static version of the model effort is identically equal to zero, and therefore disutility from effort is irrelevant. Therefore $SCEV^i$ provides a clean measure of the static gains from better consumption insurance induced by the policies against which the dynamic adverse incentive effects have to be traded off.[46]

Table 7: Aggregate Welfare Comparisons

	Static CEV^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

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

Turning now to the main object of interest, the dynamic welfare consequences in column 2 of Table 7 paint a different picture. As in static analysis, both policies improve on the laissez-faire equilibrium, and

[43] Recall that we carry out our analysis for each (*educ*)-type separately and report averages across these types. Thus in what follows Φ_0 suppresses the (policy-invariant) dependence of the initial distribution on (*educ*).

[44] Using the notation from section 4.2 for the constrained efficient allocation and for equilibrium allocations under policy i, $W(c^{SP}, e^{SP}) = V(\Phi_0)$ and $W(c^i, e^i) = \int v_0^i(h) d\Phi_0$.

[45] Recall that even the constrained social planner problem is solved for each specific (*educ*) group separately and thus also does not permit ex-ante insurance against being part of an unfavorable (*educ*)-group. We consider this restricted social planner problem because we view the results as more informative about the competitive equilibrium allocations.

[46] Similar to the dynamic we compute the static consumption-equivalent loss (relative to the competitive equilibrium) as $U(c^{CE}(1+SCEV^i)) = U(c^i)$, where $U(c)$ is the expected utility from the period 0 *consumption allocation* under the cross-sectional distribution Φ_0, and thus is determined by the static version of the model. Thus, using the notation from section 4.1, $U(c^i) = U^i(\Phi_0)$ for $i \in \{SP, NP, NW, B\}$ and $U(c^{CE}) = \int U^{CE}(h) d\Phi_0$.

the welfare gains are substantial, ranging from 3% to 5.3% 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 the better consumption insurance.[47]

But what we really want to stress is that there are crucial differences to the static analysis. It is *not* optimal to introduce a no-prior conditions law once a wage non-discrimination law is already in place. The latter policy already provides fairly effective consumption insurance, and the further reduction of incentives and the associated fall in mean consumption makes a combination of both policies suboptimal. The associated 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 combination (a wage non-discrimination law alone) leads to welfare losses relative to the constrained efficient allocation, although these losses are fairly modest, in the order of 0.53% of permanent consumption.[48] 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 modest. The last effect, however, *is* quantitatively crucial in explaining why the no-prior conditions law in isolation fares worse than the wage non-discrimination policies (and a combination of both policies, which restores efficiency in health expenditures, recall proposition 12).

6.3.2 Heterogeneity by Health Status

The welfare consequences reported in Table 7 were measured under the veil of ignorance, before workers learn their initial health level h. They mask very substantial heterogeneity in how workers feel about 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 dynamic consumption-equivalent variation measures, computed exactly as before, but now computed after the initial health status has been materialized.

Table 8: Welfare Comparison in the Dynamic Economy Conditional on Health

	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 of 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 (only the wage non-discrimination policy) and having both policies in place. In contrast, young households with *excellent* health support the lesser insurance (and lesser effort distortion) from a no-prior conditions law to a no wage discrimination legislation. Note that even for these households, however, some future insurance through the no prior conditions law is preferred to the unregulated equilibrium, even though that policy comes with approximately 1.5% lower consumption, relative to the competitive equilibrium. Note that 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% of laissez faire lifetime consumption to see the optimal policy introduced, households of excellent health would be prepared to *pay* 1% of lifetime consumption to prevent this policy innovation.

Interestingly, as a cohort ages the assessment of the desirability of the policies under consideration from different health types change. In Table 9 we display the dynamic consumption-equivalent variation measures,

[47] In addition, the NW-policy induces lower costly effort (Figure 8) but also leads to lower average consumption (Figure 13) compared to the NP-law. These last two effects roughly cancel out in terms of welfare.

[48] In contrast, an unconstrained planner that controls effort of households directly can do much better than the best policy, with a welfare gap worth 7.3% of lifetime consumption.

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 now 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 do not value insurance against health deterioration strongly, relative to the present consumption losses implied by the policies. If we translate welfare consequences of these policies into political support, as a cohort ages the opposition against far-reaching social insurance from those who have maintained very good or excellent health grows stronger.

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

7 Robustness and Sensitivity Analyses

In this section we explore the robustness of our main findings to alternative strategies of mapping the model to the data and model extensions. In subsection 7.1 we first investigate how sensitive our results are to the magnitudes of the three key model building blocks. Second, in subsection 7.2 we study how mismeasured *real* health expenditures and a mis-specified link between health expenditures and health outcomes would affect our results, and finally, in subsection 7.3 we explore the sensitivity to alternative modelling assumptions. Table 10 provides an overview of the results from these exercises.

Table 10: Dynamic Welfare Results from Sensitivity Analyses

	BM	IG	PL	QM	FX	Incomplete NW			UN
						PP	Cost	PP, Cost	
Constrained Social Planner	5.587	1.086	2.893	4.115	5.587	5.587	5.587	5.587	5.736
Competitive Equilibrium	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No Prior Conditions Law	2.904	0.710	0.564	2.170	1.920	2.904	2.904	2.904	3.071
No Wage Discrimination Law	5.055	1.045	2.718	3.697	5.010	5.409	0.010	1.652	5.218
Both Policies	2.973	0.066	0.945	1.734	2.973	5.510	-2.081	3.121	3.140

Note: BM = Benchmark Economy; IG (7.1.1) = Lower Effect of Health on Income; PL (7.1.2) = Insurance pooling as benchmark; QM (7.1.3) = Re-estimated Q with Mismeasured Effort; FX (7.2.2) = Fix Medical Expenditure; Incomplete No-Wage Discrimination (7.3.1); PP = Partial Wage Pooling, with $\tau = 0.3$; Cost $\gamma = 4.55\%$; PP + Cost = Partial Wage Pooling ($\tau = 0.8$) with 4.55% Resource Cost; UN (7.3.2) = Model with Uninsured. In PL, the CEV's are relative to the Competitive Equilibrium with insurance pooling. In the Incomplete NW Discrimination analyses with partial wage pooling ($\tau > 0$), "Both Policies" refers to implementing the no prior conditions law on top of *partial* wage pooling.

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

7.1.1 How Important is Health Status for Income

To address the concern that the causal link between health status and labor income is weaker than estimated thus far (where we already controlled for observable heterogeneity) we now conduct 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.[49] Figure 19 in appendix H.1 displays the old and the new income profiles used as targets for the estimation of the model.[50] Under the new estimation, workers with excellent health now earn about 29% more than those with fair health. This ratio is lower for young workers (ages 41 or less) at 19%. As displayed in Table 10 (second column labelled IG), since the health-induced income risk is significantly reduced, the overall welfare benefits of social insurance policies decline, but remain significant at about 1% of lifetime consumption. Crucially, our policy rankings and the conclusion that applying both policies in conjunction is suboptimal is robust to this change.

In order to get a sense whether our income gradients are empirically plausible we can also turn to an empirical literature that has estimated the earnings consequences of individual health conditions. For example, Mitchell and Butler (1986) estimate that arthritis reduces earnings by 33% and Bartel and Taubman (1979) find a negative impact of various disease ranging between 20 and 30%. We apply the same technique as in section 5.2.2 and use MEPS data to compute the expected labor earnings by health status consistent with the relative frequency of arthritis, and find that workers with excellent health earn 17% more than those with fair health. This suggests that the health impact on earnings we use in this sensitivity analysis is not implausibly large relative to the existing empirical literature.[51]

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 also because it reduces (in expectation) future health expenditures. This is the second key model element in the model. In the benchmark model estimated thus far it was assumed that health insurance premia paid by individuals reflect expenditures conditional on health status, without any pooling (this is, we estimated the model's competitive equilibrium). It is plausible to argue that the PSID data we used (from 1997 to 2003) were generated from 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 (and thus consumption) risk that households face. To evaluate the importance of this concern we now re-estimate the model under the assumption that all health insurance premium risk is already pooled. We refer to this as the insurance-pooling equilibrium.[52]

The re-estimated model needs to match the effort data with smaller benefits from being healthy. Consequently, although most parameters remain essentially unchanged, the estimated cost shifters $\gamma(h)$ fall and the $q(.)$ function becomes slightly more convex, see Table 21 in the appendix. Crucially, as Table 10 (third column labelled 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 now already provides a significant degree of consumption insurance against health risk.[53] 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[54] beneficial, even when it comes with the reductions of incentives documented in section 6.2. As before, the complete deterioration of effort incentives under a combination of both policies is strongly dominated.

[49] In order to keep average income in the economy similar to its level in the benchmark, we scale down the income ratio with respect to good health $w(h)/w(h=2)$ by half while keeping the income profile of good health the same (in levels).

[50] An alternative interpretation of this sensitivity analysis is that the original health-income gradient is estimated correctly, but that additional government insurance policies or household smoothing mechanisms we do not model here (e.g. endogenous labor supply adjustments, precautionary savings against health shocks or inter-household transfers) reduce the dependence of consumption on health status h, and by shrinking the health-income gradient we capture these additional smoothing mechanisms, although admittedly in a fairly reduced-form way.

[51] The detailed procedure and the extension of the analysis to hypertension are included in appendix H.1.

[52] Concretely, 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 now 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 (5) 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 because of the difference in the $x(\varepsilon, h)$ schedule.

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

[54] In the NW policy premium pooling is removed, (but of course implemented if both policies are present).

7.1.3 Mismeasured Effort Inputs

The third key ingredient of our model is the endogenous health effort choice that leads, in expectation, 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 (and argued that the resulting effects were empirically plausible). Here we show that the presence of unobserved (in our data set) additional effort inputs can be modelled as non-classical measurement error in the effort variable e. Assume that true effort is given by

$$e = \lambda\tilde{e} + (1-\lambda)e^* \tag{30}$$

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. If we 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 both proportional to e (see appendix H.3), and we can rewrite (30) as $e = \eta\lambda\tilde{e}$, where the random variable η has a cross-sectional distribution determined by the population distribution of the cost function parameters (ρ, ν). If that distribution is such that η is a non-negative uniform random variable, symmetrically distributed around 1 and with upper point of the support[55] 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 $\lambda = 1$, the distribution of η is degenerate, and observed effort is a perfect proxy for true effort, as assumed thus far.

When we assume a measurement error of $\lambda = 0.75$ (the midpoint in the interval between no, and maximal permissable error) in the re-estimated Q (displayed in Figure 20 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.

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 cross-sectional dispersion in nominal health expenditures observed in the data partially stems from dispersion of prices different individuals pay 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 healthcare. 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 real services $x(\varepsilon, h)$ are contracted upon. Then p and x are *uncorrelated* and health insurance 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.

[55] 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.

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 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[56] 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. To assess this concern consider the other extreme that health expenditures x are pure waste, or yield direct utility separable from consumption. Since the estimation of the competitive equilibrium would still match the same wage and health expenditure data, both the income-health 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 (as fully characterized by the expenditure cutoffs). 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.[57] This finding suggests that the exact motivation for spending on health (productivity-enhancing in the benchmark model) 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.

7.3 Sensitivity Analysis III: Robustness to Alternative Modelling Assumptions

7.3.1 Resource Cost and Limited Effectiveness of No Wage Discrimination

Thus far we have analyzed idealized versions of the ADA in the U.S. Now we investigate the impact of this policy under the more empirically realistic assumption that the ADA eliminates some, but not all health related income variation, and its implementation requires a real resource cost. We parameterize by γ the resource cost and by τ the effectiveness of the wage non-discrimination law. The labor income an individual earns is now given by the convex combination (with weight τ) of individual, health-contingent wages and average wages $\tau \times w(h) + (1 - \tau) \times (1 - \gamma) \times w$. Consumption is given by the difference between the wage and the health insurance premium.[58] In the benchmark the policy was fully effective and costless $\tau = \gamma = 0$.

In figures 14 and 15 we plot, against the effectiveness τ of the wage non-discrimination law, aggregate welfare under the various policies. The left panel has no resource cost, $\gamma = 0\%$, and the right panel imposes a resource cost of $\gamma = 4.55\%$, the midpoint of the estimates in Acemoglu and Angrist (2001).[59] The benchmark results are annotated 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 (dashed line with $\tau = 0$) or the unregulated equilibrium (partial no wage with $\tau = 1$). Abstracting from a resource cost (focusing on figure 14), as the effectiveness of the policy declines (as τ increases along the x-axis), welfare in a policy environment that includes wage compression initially *rises* as the lower consumption insurance raises effort incentives. As wage pooling becomes highly ineffective (for $\tau > 30\%$) the demand for more

[56] With a price variance of $\sigma_p^2 = 0.494^2$ the model fits the expenditure variance in the data perfectly.

[57] Since the cutoffs are identical in the CE, the constrained efficient allocation and under both policies, there is no change at all 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 nontrivial effect on the welfare results.

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

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

consumption insurance and the fact that implementing both policies jointly does not lead to a collapse in effort incentives induces a policy preference for this option, relative to the very imperfect wage pooling policy.

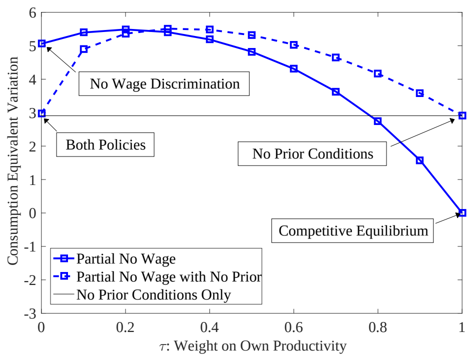


Figure 14: No Resource Cost ($\gamma = 0$)

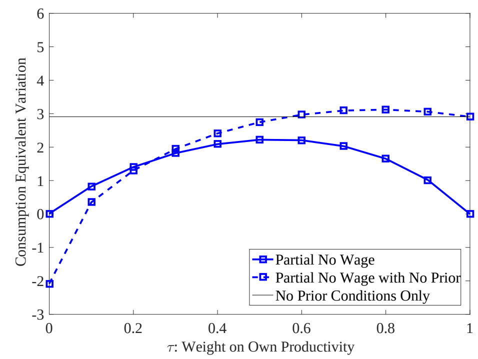


Figure 15: Positive Resource Cost ($\gamma = 4.55\%$)

Figure 15 displays the results if the wage non-discrimination law is costly. The two key observations are that now premium pooling in the health insurance market dominates costly wage pooling in the labor market and that introducing both policies jointly is only optimal if wage pooling is so imperfect (high τ) that significant incentives for maintaining health efforts continue to be maintained.[60]

The results in this section confirm and extend the main point of this paper. Empirically, there appear to be significant income and health expenditure risks associated with health of the working lifespan. At the same time, individuals have some stochastic control over their health transitions. Implementing insurance against health risk is therefore valuable but it comes at a substantial incentive cost. We therefore find that the optimal amount of consumption insurance is partial, with non-trivial remaining risk to maintain incentives. With idealized versions of our policies, this leads to wage non-discrimination law being optimal. With implementation costs and/or only partial wage insurance what policy combination is optimal not surprisingly 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: full consumption insurance through pure versions of both policies is therefore robustly suboptimal.

7.3.2 Modelling Uninsured Households

The purpose of this subsection is to demonstrate that the positive and normative implications of our model remains 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-endogenous 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 characterized in proposition 8 (see equation (10)) offered by a production firm that takes health expenditures of workers as given. As we showed in that proposition, this wage contract induces individuals (in this case the uninsured)

[60] Note that as τ increases and the wage pooling policy becomes less effective, the total cost of the policy (which by assumption is proportional to $(1-\tau)w$) declines. Had we assumed a fixed cost of the policy the no-prior conditions legislation only would have been preferred to both policies jointly, for all τ.

[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 condition, mandatory health insurance legislation. The lower bound is infinitely negative, if the policy forces individuals with infinite fixed costs to nevertheless buy insurance.

[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.

to spend x out-of-pocket according to the competitive schedule. 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 of the model continuation utility now involves taking 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.[63] The ε induced consumption risk is non-trivial and raises the attractiveness of the no-prior conditions law, but not nearly enough to change the policy rankings reported in the benchmark economy without uninsured.[64]

To conclude this section, we want to offer a justification for why we do not make the model analyzed in this section of the paper our benchmark? In the 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 observing individuals without insurance, it must be made inefficient for some of them to purchase health insurance. 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 make 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 is not informative about the fixed cost. However, our *normative* analysis of the no-prior conditions law is highly 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 want to stress again explicitly, that our analysis should not be interpreted as a comprehensive investigations into all aspect of the ACA (and specifically not the implied attempt to reduce the share of the uninsured) but rather one of its key components, outlawing the dependence of health insurance premia on prior health conditions.

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 both a no-wage discrimination intervention in the labor market, in combination 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, are large. Even though both policies improve upon the laissez-faire equilibrium, implementing them *jointly* and providing full consumption insurance against health-related risk is suboptimal. More broadly, our paper therefore shows that a reliable policy analysis of health insurance reforms on one side and labor market reforms cannot be conducted separately, since their interaction might deliver less favorable welfare results than suggested by an isolated analysis of these policies.

Based on our findings, there are several extensions we view 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, although the positive effect of health h on the continuation utility after retirement partially captures this in our model. Moreover, 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. Future work has to uncover whether such an extension of the model also affects, quantitatively or even qualitatively, our conclusions about the *relative* desirability of these policies.

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

[64] Wage insurance is also more valuable for the uninsured relative to the benchmark because of their health spending risk.

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 (2013), "Equilibrium Labor Market Search and Health Insurance Reform." *NBER Working Paper No. 18698.*

[3] Ales, L., R. Hosseini, and L. Jones (2012), "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 Medicaids Effects on Clinical Outcomes." *New England Journal of Medicine*, 368(18).

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

[7] Battacharya, 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.

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

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

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

[11] Colman, G. and D. Dave (2013), "Physical Activity and Health," *NBER Working Paper* 18858.

[12] 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.

[13] DeLeire, T. (2000) "The Wage and Employment Effects of the Americans with Disabilities Act." *Journal of Human Resources,* Volume 35, No. 4, 693-715.

[14] DeLeire, T. (2001) "Changes in Wage Discrimination against People with Disabilities: 1984-93." *Journal of Human Resources,* Volume 35, No. 1, 144-158.

[15] De Nardi, M., E. French and J. Jones (2015), "Medicaid Insurance in Old Age," *American Economic Review*, forthcoming.

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

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

[18] 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.

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

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

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

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

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

[24] 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.

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

[26] 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.

[27] 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.

[28] Kopecky, K. and T. Koreshkova (2015), "Old, Sick, Alone and Poor: A Welfare Analysis of Old-Age Social Insurance Programs," *Review of Economic Studies*.

[29] 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.

[30] Madrian, B. (1994), "Employment-Based Health Insurance and Job Mobility: Is There Evidence of Job-Lock?," *Quarterly Journal of Economics*, 109, 27-54.

[31] 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.

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

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

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

Appendix for Referees and Online Publication

A Proofs of Propositions

Proposition 5

Proof. Since exercise does not carry any benefits in the static model, trivially $e^{SP}=0$. Attaching 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

$$-F_2(h,\bar{\varepsilon}^{SP}(h)) = 1. \tag{31}$$

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 (31). ∎

Proposition 6

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

$$\begin{aligned} u'(w(h)-P(h)) &= \lambda(h) = -\mu(h) && (32)\\ \lambda(h)F_2(h,-x(0,h)) &\leq \mu(h) && (33)\\ &= \text{ if } x(0,h) > 0 &&\\ \lambda(h)F_2(h,\varepsilon - x(\varepsilon,h)) &\leq \mu(h) && (34)\\ &= \text{ if } x(\varepsilon,h) > 0 && \end{aligned}$$

Thus off corners we have

$$F_2(h,\hat{\varepsilon} - x(\hat{\varepsilon},h)) = F_2(h,\varepsilon - x(\varepsilon,h)) = K \tag{35}$$

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

$$x^{CE}(\varepsilon,h) = \max\left[0,\varepsilon - \bar{\varepsilon}^{CE}(h)\right]. \tag{36}$$

Plugging (36) into (34) and evaluating it at $\varepsilon = \bar{\varepsilon}^{CE}(h)$ yields

$$\lambda(h)F_2(h,\bar{\varepsilon}^{CE}(h)) = \mu(h). \tag{37}$$

Using this result in the second part of (32) 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 8

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 (10). 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 (10) 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

$$-F_2(h, \varepsilon - x(\varepsilon, h)) = \frac{\int u'\left[\boldsymbol{w}(h, \varepsilon - x(\varepsilon, h)) - P(h)\right] f(\varepsilon) d\varepsilon}{u'\left[\boldsymbol{w}(h, \tilde{\varepsilon} - x(\tilde{\varepsilon})) - P(h)\right]}. \tag{38}$$

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 (38) 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

$$F(h, \varepsilon - x) - w^{CE}(h) = \begin{cases} F(h, \varepsilon) - w^{CE}(h) & \text{if } \varepsilon < \bar{\varepsilon}^{CE}(h) \\ F(h, \bar{\varepsilon}^{CE}(h)) - w^{CE}(h) & \text{otherwise} \end{cases}. \tag{39}$$

Now we show that under the proposed separation of wage and health insurance contracts the production firm does not need to know whether the 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 is 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 the worker buys insurance for health shocks $\varepsilon \geq \bar{\varepsilon}^{CE}(h)$ or not, vacating the need for the firm to verify whether the worker has purchased health insurance elsewhere or not. ■

Proposition 9

Proof. Let Lagrange multipliers to equations (5) 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

$$u'(w(h)-P)[-F_2(h,\bar{\varepsilon}^{NP}(h))] = \sum_h u'(w(h)-P)\Phi(h). \tag{40}$$

Moreover, let us take the derivative of (40) 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 10

Proof. From (11), P we immediately obtain

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

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

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

$$-F_2(h_L,\bar{\varepsilon}^{NP}(h_L)) > 1 > -F_2(h_H,\bar{\varepsilon}^{NP}(h_H)), \tag{41}$$

i.e.

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

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 (44). This result, in combination with (41) 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 (11). ■

Proposition 14

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 is 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 (13), and given these health distributions wages and health insurance premia are given by (20) and (22). ■

B Further Analysis of the No-Wage Discrimination Case

B.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 assume 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

$$w(h) = g(h)F(h,0) + (1-g(h))\int_0^{\bar{\varepsilon}} f(\varepsilon)[F(h,\varepsilon - x^{NP}(\varepsilon,h))]d\varepsilon, \tag{42}$$

holds and they are fully employed, or $w(h) - P(h) = 0$. On the flip side, there will be excess demand for workers whose expected production is more than $w(h)$, they will therefore find it optimal to either lower their insurance, and in the limit as $\varepsilon \to 0$ either (42) holds they or set $x(e,h) = 0$ if they end up at 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.

B.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

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

since health type 0 will have the lowest wage in equilibrium, and where C is the penalty parameter and $n(h)$ is 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 workers low health status workers raising their productivity by over-insuring themselves against health risks and high health status workers 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

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

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

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

is positive and less that 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 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 16 *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.*

B.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 17 *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 the 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

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

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

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

This condition implies that a firm will hire more that 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 that are not fully employed in equilibrium would have excess members who would be happy to be hired 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 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.

C 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 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 $\bar{\varepsilon}^{CE}(h)$, (9), 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 (9), implies that

$$\frac{dw^{CE}(h)}{dh} = g'(h)\left[\begin{array}{c} F(h,0)-\int_0^{\bar{\varepsilon}^{CE}(h)} f(\varepsilon)F(h,\varepsilon-x(\varepsilon,h))d\varepsilon \\ -\int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F(h,\bar{\varepsilon}^{CE}(h))d\varepsilon \end{array}\right] \quad (43)$$

$$+g(h)F_1(h,0)+(1-g(h))\int_0^{\bar{\varepsilon}^{CE}(h)} f(\varepsilon)F_1(h,\varepsilon-x(\varepsilon,h))d\varepsilon$$

$$+(1-g(h))\int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F_1(h,\bar{\varepsilon}^{CE}(h))d\varepsilon$$

$$-(1-g(h))\int_{\bar{\varepsilon}^{CE}(h)}^{\bar{\varepsilon}} f(\varepsilon)F_2(h,\bar{\varepsilon}^{CE}(h))\frac{F_{12}(h,\bar{\varepsilon}^{CE}(h))}{F_{22}(h,\bar{\varepsilon}^{CE}(h))}d\varepsilon.$$

All of the terms in (43) 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, then wages will vary positive with health status. Note that this is trivially implied if the direct effect of the change in health status offsets the spillover, or

$$F_1(h,\bar{\varepsilon}^{CE}(h)) - F_2(h,\bar{\varepsilon}^{CE}(h))\frac{F_{12}(h,\bar{\varepsilon}^{CE}(h))}{F_{22}(h,\bar{\varepsilon}^{CE}(h))} > 0. \quad (44)$$

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 model elements. We summarize our results in the following proposition:

Proposition 18 *The competitive wage is increasing in h if (43) is positive.*

D Computation of the Social Planner Problem

The social planner problem in section 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 recursive problem of the planner has as 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 assure that the aggregate resource constraint is satisfied in every period t. Thus the problem we solve is

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

$s.t.$

$$V_t(h) = u(c_t(h)) - q(e_t(h)) + \beta \sum_{h'} Q(h';h,e_t(h))v_{t+1}(h') \quad (45)$$

$$q'(e_t(h)) = \beta \sum_{h'} \frac{\partial Q(h';h,e_t(h))}{\partial e_t(h)} v_{t+1}(h'), \quad (46)$$

$$\Phi_{t+1}(h') = \sum_h Q(h';h,e_t(h))\Phi_t(h). \quad (47)$$

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 is κ the harder might the optimization problem be solved. This suggests the following algorithm

Algorithm 19 *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 initial guess. 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*
$$\left\|c^{n+1} - c^n\right\| < \varepsilon.$$

E 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 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 the briefly discuss how it is modified for the other policy regimes:

Algorithm 20 1. *Guess a sequence $\{Eu'_t, P_t\}_{t=0}^T$.*

2. *Given the guess use equations* (23), (25), (21) *to determine health cutoffs and wages $\{\bar{\varepsilon}_t^{NP}(h), w_t(h)\}$.*

3. *Given $\{w_t(h), P_t\}$, solve the household dynamic programming problem* (17) *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* (13).

5. *Obtain a new sequence $\{Eu_t'^{new}, P_t^{new}\}_{t=0}^T$ from* (20) *and* (21).

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 (23), (26), (19) 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 (19) and (22). With both policies, equation (22) replaces (21) in all expressions.

F Details for Data and Estimation

F.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.

F.2 Descriptive Statistics

Before we proceed to descriptive statistics of the PSID data, we summarize, in Table 11, the mapping between variables in our model and data.

Table 11: 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 average of reported medical expenditure and labor income over 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 12: 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

Table 12 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 expenditures. The reported labor income is the average labor income over three survey years.

Medical expenditure data is from MEPS and are averaged over six survey years (1997 through 2002). While PSID only reports medical expenditures at the household level, MEPS contains individual-level medical expenditure, which is the relevant statistic for us. In both surveys, health status are reported as Excellent, Very Good, Good, Fair, and Poor, but we group Poor with Fair to ensure 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, a value of 1.

F.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 13 summarizes the transition matrix for health status from the raw, untreated PSID data.

Table 13: 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 are:

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

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.

Effects of Exercise on Health Status: Evidence from the Empirical Literature In order to compare our estimated effects of effort on health updates, we now map data on health conditions data to the health status h data we use. Table 20 documents the prevalence of specific diseases and smoking behavior by health status, based on MEPS 2000 and 2001 data.

Using this data, we can compare our estimated effort effects on health outcomes with the empirical study by Colman and Dave (2013). Colman and Dave (2013) finds that physical exercise reduces risk factors for health diseases, and has a lagged effect that endures over time. They report that high level of recreational exercise reduces the probability of having hypertension by between 5.1 and 8.4 percentage points (significant at 0.1%) depending on intensity, using the lagged outcome model. Similarly, the effect ranges between 1.8 and 2.2 percentage points (significant at 1%) for diabetes, and 1.9 and 2.4 percentage points (significant at 1%) for heart disease. They also report effects of non-recreational exercises and estimates from the fixed effects model.

In order to compare our estimates of effort on health transitions to specific disease incidence in the main text, we specifically conduct the following thought experiment:

Figure 16: Transition for Low Education

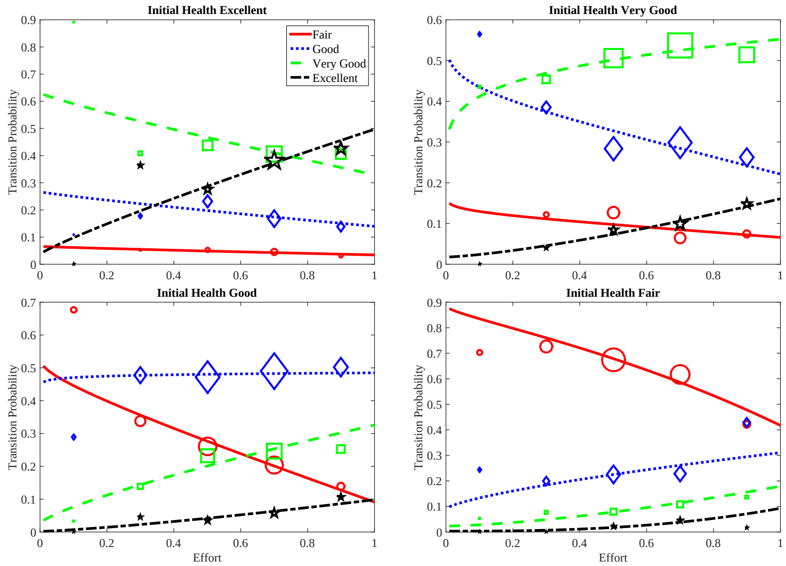
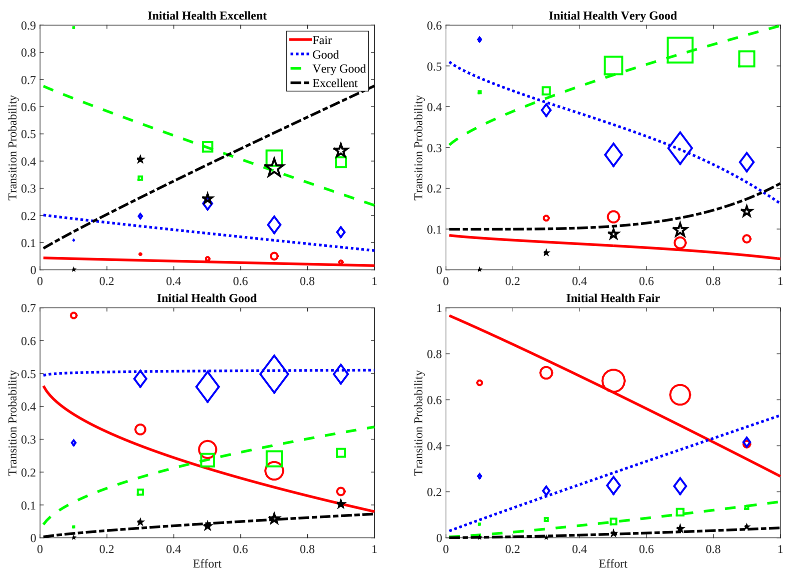


Figure 17: Transition for High Education



(a) Construct a mapping between health status and probability of disease (from MEPS, Table 20)

(b) Divide the sample into initial health (4) - exercise (3) groups (since Colman and Dave (2013) also use three exercise levels)

(c) Using our benchmark estimated health transition function, for each initial health status and for each exercise level, we calculate the person's probability of having a disease in the future by using the formula:

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

(d) Obtain the implied changes in probability of having a disease (through changes in health status) for different exercise groups, and compare the magnitude to that reported by Colman and Dave (2013).

Table 14: Disease Incidence and Smoking by Health Status[1]

	Diabetes	Hypertension	Heart Disease[2]
Fair	0.2205	0.4897	0.2307
Good	0.0856	0.3040	0.0879
Very Good	0.0339	0.1927	0.0526
Excellent	0.0120	0.1094	0.0307
Total	0.0640	0.2377	0.0761
Observations	32,312	32,865	38,601

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

[2] An individual is reported to have a heart disease if the person had ever been diagnosed with a coronary heart disease; angina or angina pectoris; heart attack or myocardial infarction; or another kind of heart disease or condition; or a stroke or transient ischemic attack.

Table 15: Effect of Exercise on Incidence of Diseases from the Estimated Health Transition

Disease	Diabetes			Hypertension			Heart Disease		
Exercise Level	Low	Middle	High	Low	Middle	High	Low	Middle	High
Fair	0.175	0.160	0.144	0.423	0.400	0.376	0.184	0.169	0.153
Good	0.106	0.096	0.087	0.321	0.303	0.289	0.115	0.105	0.096
Very Good	0.064	0.060	0.055	0.247	0.238	0.228	0.077	0.073	0.069
Excellent	0.041	0.038	0.035	0.191	0.182	0.172	0.057	0.054	0.051

Table 15 summarizes our findings. We observe that, depending on the specific health condition, our health transition function, a 2-3 percentage points decline in the incidence of disease (in the future) from higher effort, which is consistent with the empirical estimates from Colman and Dave (2013).

F.4 Generating 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 average of $\ln \tilde{w}$ is equal to the average of $\ln w$ from 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 18.

F.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, $\mu_z(t, educ, h)$ are exogenously given. We construct these parameters using the similar procedures as 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 16.

Table 16: 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

As for our ε-shock expenditure profiles, we use the smoothed moments as our targets, which are presented in Table 18.

The remaining parameters for health shock distribution are health-dependent probabilities of not getting health shocks $\tilde{g}(h)$, the parameter that governs age-effect on probability α_g, mean and variance of ε-shock distribution μ_ε and σ^2_ε, 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.

F.6 Estimation Results

Table 17 summarizes the value of parameters that were estimated inside the model using efficient GMM. Table 18 contains the details of the model fit with respect to income and medical expenditure profiles.

G Additional Quantitative Results

Tables 19 presents the static and dynamic consumption equivalent variations for each education group.

H Other Robustness and Sensitivity Analysis: Details

H.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 lower income gradient of health and evaluate the policy consequences. In Figure 19, 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

Table 17: 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

Table 18: 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

Figure 18: Distribution of (log) Medical Expenditure

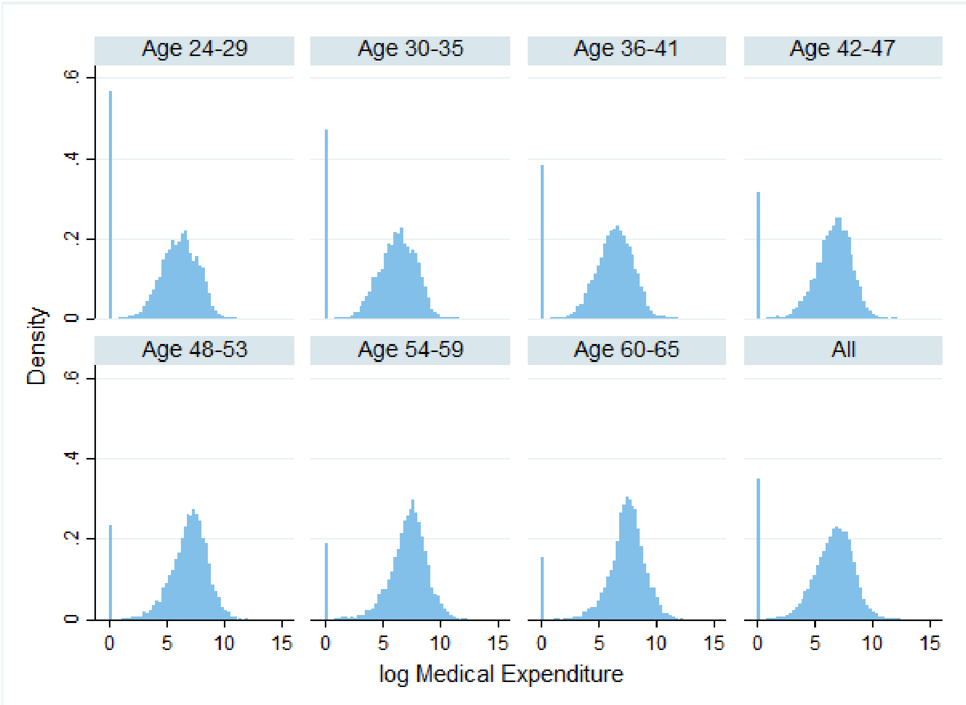


Table 19: Welfare Comparisons by Education

	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

and our lower gradient estimation for Fair (normalized to 1), Good, Very Good, and Excellent health are 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.

We would also like to check the plausibility of our labor income profiles by comparing them to the estimates from the empirical literature. Bartel and Taubman (1979) finds significant negative effects of health conditions that range between 20 and 30% reductions in earnings (as lower bound), and Mitchell and Butler (1986) estimates a 33% decline in the wages of workers with arthritis. Normalizing the income of

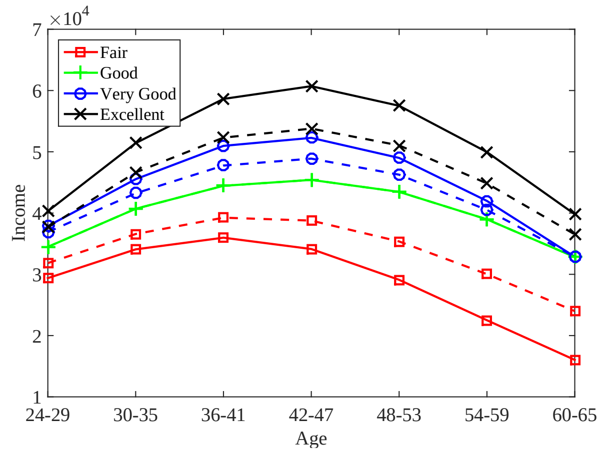


Figure 19: Labor Income Profiles: Benchmark (Solid) vs. Lower Health-Income Gradient (Dotted)

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 20, we document the prevalence of hypertension and arthritis and implied earnings by health status, using 25% income loss for hypertension (a mid-point from Mitchell and Butler) and 33% for arthritis. We find that they lead to about 11 to 17% loss in expected earnings, lower than but still roughly similar to the income ratios used in this sensitivity analysis, especially for young workers.

Table 20: Disease Incidence and Implied Earnings by Health Status[1]

	Hypertension		Arthritis	
	Probability	Earnings Ratio	Probability	Earnings Ratio
Fair	0.490	1.000	0.522	1.000
Good	0.304	1.053	0.293	1.091
Very Good	0.193	1.085	0.120	1.130
Excellent	0.109	1.108	0.105	1.166

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

H.2 Insurance Pooling as the Benchmark

Table 21 summarizes the key parameter changes induced by switching the estimated economy to the partial no prior conditions economy.

Table 21: 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

H.3 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 (30) yields as optimality condition

$$\frac{\tilde{e}}{e^*} = \left[\frac{\lambda(1-\rho)}{(1-\lambda)\rho}\right]^{\frac{1}{\nu-1}} := \kappa(\rho, \nu; \lambda). \tag{48}$$

Using this equation in 30 yields as optimal solution $\tilde{e} = \Gamma_1(\rho, \nu; \lambda)e$ and $e^* = \Gamma_2(\rho, \nu, \lambda)e$, where

$$\Gamma_1(\rho, \nu; \lambda) = \frac{\kappa(\rho,\nu;\lambda)}{\lambda\kappa(\rho,\nu;\lambda)+1-\lambda} \tag{49}$$

$$\Gamma_2(\rho, \nu; \lambda) = \frac{1}{\lambda\kappa(\rho,\nu;\lambda)+1-\lambda} \tag{50}$$

Plugging these solutions into 30 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 utility cost of providing true effort e. This requires

$$\Psi(\chi) \left[\rho\Gamma_1^\nu + (1-\rho)\Gamma_2^\nu\right]^{\frac{1}{\nu}} = 1. \tag{51}$$

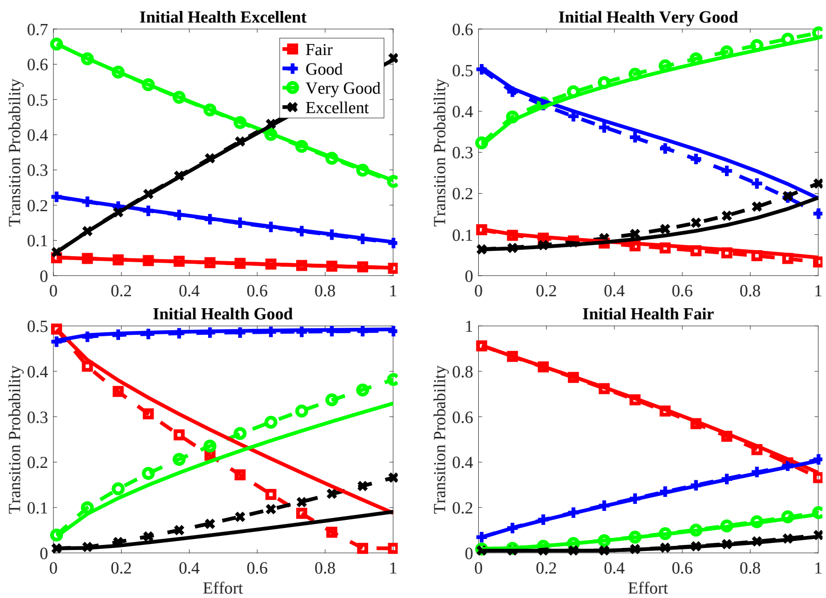
Note that under the assumptions on the cost function C, given the observation $\tilde{e}$ and given an assumed distribution for χ, the other aspects of the model are not needed to infer the distribution of e^*. This continues to permit us to estimate Q outside the model despite the fact that $\tilde{e}, e^*$ 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 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 is $\tilde{e}$ 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 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

$$\max_\theta \sum_i \int_{2\lambda-1}^{1} \log\left[\frac{Q(h_i'|h_i, \eta\lambda\tilde{e}_i, \theta)}{(2-2\lambda)}\right] d\eta \tag{52}$$

where i indexes the household effort and health status observations. For our application in the main text we set $\lambda = 0.75$. Figure 20 display our new estimates of Q (weighted averages across education groups), against the estimated transitions in the benchmark.

Figure 20: Transition Averaged over Education



Note: Solid lines represent our benchmark specification without measurement error. The dotted lines with markers represent the estimated transition functions with measurement error and $\lambda = 0.75$.