CEWP 25-02

On the Over-Provision of Medical Insurance

Afrasiab Mirza
University of Birmingham

Eric Stephens
Carleton University

March 10, 2025

CARLETON ECONOMICS WORKING PAPERS

Department of Economics

1125 Colonel By Drive
Ottawa, Ontario, Canada
K1S 5B6

On the Over-Provision of Medical Insurance

Afrasiab Mirza
Department of Economics
University of Birmingham
a.mirza@bham.ac.uk

Eric Stephens
Department of Economics
Carleton University
eric.stephens@carleton.ca

December 22, 2024

Abstract

This paper considers the general equilibrium implications of moral hazard in private health insurance markets. We show that the structure of standard contracts gives rise to a pecuniary externality whereby individuals ignore the impact of their insurance purchases on the future price of care. At the equilibrium, individuals over-insure against health expenditure risk, and over-spend on medical services while facing an excessive price of care. Reducing insurance coverage at the margin can mitigate the externality by exerting downward pressure on prices, thereby raising welfare.

JEL codes: D52; I11; I13; I18

1 Introduction

This paper develops a general equilibrium model of health insurance with linear contracts similar to those typically observed in practice. We show that the presence of moral hazard implies the existence of a pecuniary externality that arises from the failure of individuals to account for the effect that their insurance purchases have on the future price of care. We show that this standard contract structure causes households to over insure against health expenditure risk, over spend on medical services when ill, and results in a price of care that is excessively high.

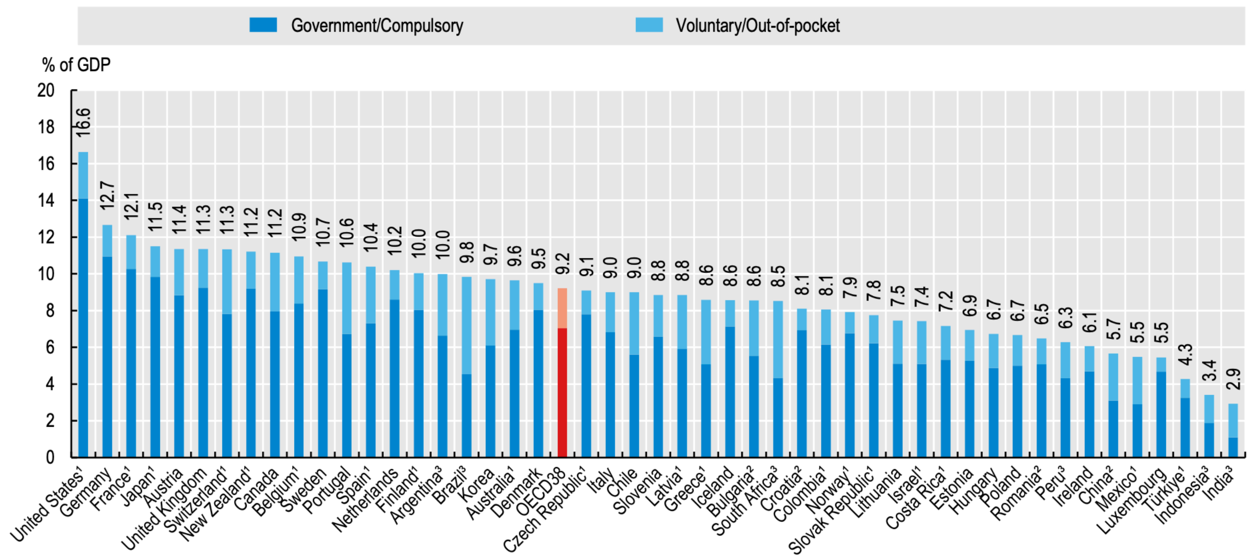
That linear contracts generate a sort of moral hazard, in which individuals may overspend on health care at the margin has been well established since Pauly (1968). The welfare implications of this are unclear however, since a sub-optimal contract implies that any efficiency arguments are applicable only to a constrained, or second-best allocation. In contrast, we establish, under standard testable assumptions on preferences and the production of medical care, that the presence of moral hazard implies an equilibrium that is generally *constrained inefficient*. Thereby offering a rationale for public intervention in health care.[1]

Spending on health care constitutes a significant portion of GDP in nearly all OECD countries (see Figure 1). While social health insurance represents the lion's share of health financing in most countries, private health insurance accounts for nearly 10% of health care spending across the OECD (see Figure 2). For example, private insurance in the United States accounts for a third of all health spending, nearly half in Switzerland, and around 60% in the Netherlands. Furthermore, in the United States the introduction of the Affordable Care Act in 2014 has resulted in substantial growth in private individual health insurance with direct-purchase plans covering an estimated 46 million individuals (13.9% of the population) in 2022.[2]

Health insurance contracts are often annual and typically consist of different prices over

[1] The externality we identify is generic and will be present in all forms of private health care provision.

[2] OECD Health Statistics 2021, https://doi.org/10.1787/health-data-en, and Congressional Research Report F10830.



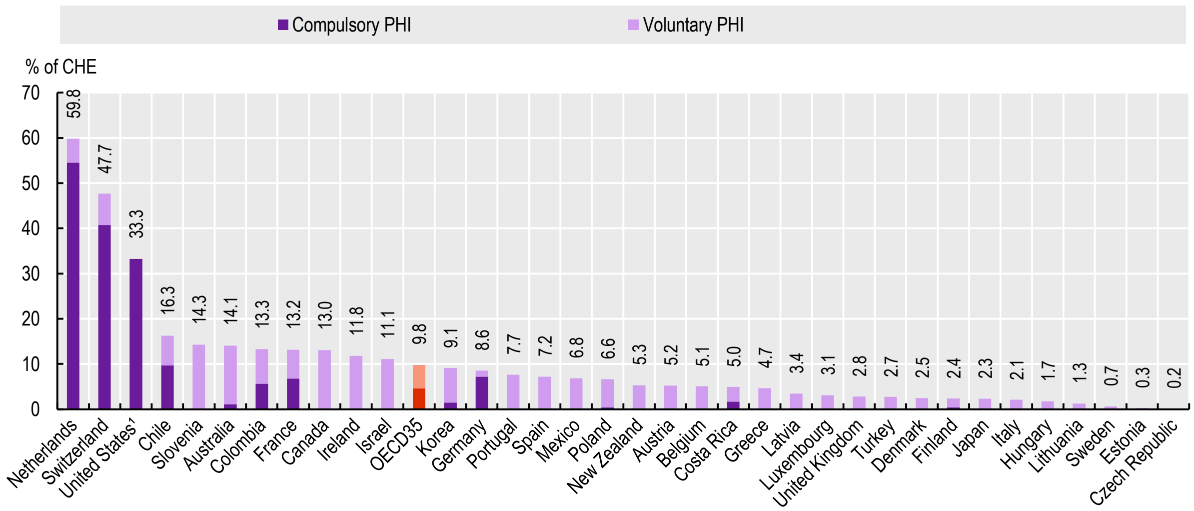
1. OECD estimate for 2022. 2. 2021 data. 3. 2020 data.
Source: OECD Health Statistics 2023; WHO Global Health Expenditure Database.

Figure 1: Health expenditure as a % of GDP.

three regions of expenditure. There is a deductible range, in which all medical expenses are out of pocket. This is followed by a copay or coinsurance range, in which individuals are responsible for a fraction of total costs. For example, in the copay region the insured may pay 10% of total costs. Finally, there may be a stop-loss provision or catastrophic coverage amount, which is an upper bound on the out of pocket expense. This type of contract is illustrated in Figure 3.

The contract structure that is illustrated in Figure 3 is ubiquitous because of the complex nature of individual health, which makes it difficult to specify all (medical and other) contingencies at the outset. Moreover, certain outcomes, like the severity of an illness or effort in finding low cost care, may not be verifiable by third parties and thus cannot be written into an enforceable contract. Since the price of care is distorted at the margin, it has been well-established in the literature that these types of health-insurance contracts give rise to an ex-post moral hazard problem.[3] Moral hazard here referring to the phenomenon whereby individuals spend more on care the lower their copay. We use this terminology

[3]For comprehensive surveys see Cutler and Zeckhauser (2000), Zweifel and Manning (2000) and Einav and Finkelstein (2018).



Note: Total private health insurance spending is defined as the sum of spending by compulsory private health insurance schemes and voluntary private health insurance schemes. CHE stands for current health expenditure.

1. Spending by private health insurance cannot be distinguished between compulsory and voluntary. Since the introduction of the individual mandate to purchase health insurance 2014 as part of the Affordable Care Act, the majority is considered as compulsory.

Figure 2: Private health insurance spending on care as a % of health expenditure.

because it is common in this literature, but note that it does not necessarily refer to an explicit hidden-action problem.

Our focus in this paper is not on an optimal contracting arrangement, but rather on the efficiency implications of the simple piece-wise linear contract structure that is usually observed in practice (as well as much of the literature). To analyze the general equilibrium implications of this typical (inefficient) contract structure we develop a dynamic model of health insurance and care provision with two periods. In the first period, individuals purchase insurance contracts to help cover health expenditure in the second period, where they face the risk of illness and may need to spend on care. The insurers are competitive and the optimal linear contract is actuarially fair and features a certain level of subsidization captured by a copay arrangement. On the care provision side of the market, our model features a large number of suppliers which are characterized by a simple production function, and an alternative technology. They invest in increasing the supply of medical care till the marginal return matches that from their alternative investment option.

Often the debate over healthcare provision and insurance ignores the optimal individual

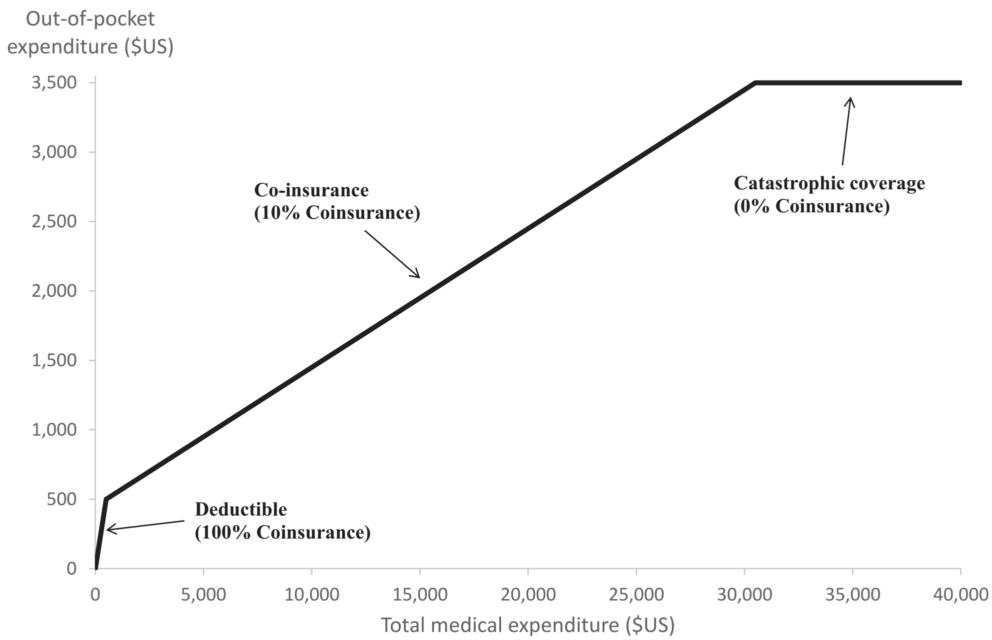


Figure 3: This illustration is taken directly from Einav and Finkelstein (2018), in which it is labeled Figure 1. It shows a stylized annual health insurance contract, illustrating the mapping the contract creates from total medical spending to out of pocket medical spending. The x-axis shows total medical spending for the year and the y-axis shows the out-of-pocket medical spending for the year.

risk level, despite the dramatic rise in insurance coverage and aggregate health spending in recent years. When contracts are linear, risk ought not be zero as would be the case in a first best scenario. Thus, despite the fact that individuals' are generally under-insured relative to the first best, it is not clear whether increasing insurance would be beneficial. We show that insurance coverage will be excessive due to aggregate pricing effects that arise in a general equilibrium. This is true when the demand for care is downward sloping and the supply is upward sloping. There is significant evidence in the literature to suggest that demand is indeed sensitive to the price of care, for a discussion see Einav and Finkelstein (2018). On the other hand, there is little existing evidence regarding supply, and the sensitivity to price remains an open empirical question.

Since insurance is excessive in our environment, we show that ex-ante limits on insurance coverage can improve welfare. It is important to note that the pricing implications of the contract structure that we highlight arise regardless of the specific optimal copay level, i.e., irrespective of the conditions that pin down the optimal copay in a partial equilibrium analysis. For example, the optimal copay could be a function of any number of factors that have been discussed in the literature such as the elasticity of demand, the tax system (e.g. Feldstein and Friedman (1977)), or information on the effectiveness of medical treatment (e.g. Chernew, Rosen, and Fendrick (2007), and Pauly and Blavin (2008)) or even behavioral concerns (e.g. Baicker, Mullainathan, and Schwartzstein (2015)).

Related Literature

Our analysis is motivated by the tremendous growth in spending on health and the inefficient nature of insurance contracts. Much of the relevant previous theoretical research has tended to focus on partial equilibrium and the theory of the second best; namely the interaction between various frictions. For example (Gaynor, Haas-Wilson, and Vogt (2000), and Wigger and Anlauf (2007)) consider the interaction between moral hazard and the price of care when prices are inefficiently high due to supplier market power. We abstract from supplier market power and focus on a somewhat more fundamental issue; that is the implications of incomplete markets for risk on the price of care in a competitive environment where suppliers are price takers. It is clear that the inefficiency we focus on is different from these papers, since they assume a constant marginal cost of production, which in our model precludes any pricing inefficiency.

The theoretical literature has shown that constrained inefficiency is a generic phenomenon when markets are incomplete.[4] In our setting, moral hazard implies that market are never complete. As a result, increasing insurance (further completing markets) in such a setting, may reduce welfare. This is different from the type of inefficiency highlighted in Pauly (1968) and the related existing literature on health insurance. In our model, the allocation is not only inefficient relative to the first best, but also the second-best, since the choice of the

[4] See Hart (1975), Greenwald and Stiglitz (1986), and Geanakoplos and Polemarchakis (1986).

insurance copay does not take into account general equilibrium implications. The resulting price effects have previously been overlooked, since the supply side of the market has been largely ignored.

Feldstein (1973), which was subsequently updated by Feldman and Dowd (1991), also argues that insurance is excessive and empirically estimates the extent of overspending. In contrast, we make a theoretical case that insurance is excessive relative to the second-best. In other words that there is scope for intervention due to the presence of a pecuniary externality.

Also relevant is Finkelstein (2007), which provides evidence on the importance of accounting for general equilibrium effects when considering the dramatic rise in health spending in the US over the previous 50 years. Her results suggest that incorporating general equilibrium effects can lead to much greater impacts of expanding health insurance than extrapolation from micro level studies would suggest. This sentiment is echoed in Gruber and Levy (2009), which discusses many of the major issues facing the US. The authors conclude that "the real problem facing the health insurance system in the United States is not so much the risk of high spending by individual households as the system wide risk of increasing aggregate spending." Our model provides a tractable framework that embeds the standard moral hazard problem and allows for a welfare analysis of changing the aggregate level of health insurance in a general equilibrium context.

In the next section we present the model, while section 3 characterizes optimal behavior and the competitive equilibrium. Section 4 examines the efficiency of the market allocation, and Section 5 concludes.

2 Model

Consider a two-period economy, where time is denoted by $t = 0, 1$. There is a measure one of ex-ante identical individuals, insurance firms and health care providers, each of which are described in turn below.

2.1 Individuals

Individuals have preferences over consumption c_t, and health h_t, in each period that are state-dependent. We denote by $u(c, h)$ and $v(c, h)$ the utility functions of agents that are healthy and ill respectively, which we assume are increasing (in consumption and health), and concave. We also assume that individuals own an equal share of any profits made by firms. Risk-less savings at time 0 are denoted b, which provide a return of R at $t = 1$.

Each individual receives an endowment of income y_t at the start of each period, and is initially endowed with h_0 units of health. During period 1, they may experience a bout of illness that causes a decline $S \in \mathbb{R}_+$ in health, which occurs with probability π. Individuals can offset the effects of illness by purchasing health care m at a unit price p. Thus an individual who experiences a bout of illness during period 1, and consumes m units of care, has a level of health $h_1 = h_0 - S + m$ at the end of period 1.

2.2 Insurance Firms

The cost of care may be offset by medical coverage that pays for a fraction of the costs incurred by individuals, which we denote by ϕ. Insurance may be purchased at the start of $t = 0$, in which case an upfront premium q is paid. Furthermore, we assume that there is perfect competition in the insurance market so that insurers break even on each contract.[5] Note that we ignore the deductible and stop-loss ranges of expenditure in the typical contract outlined in Figure 3. This is done for exposition since all that is required for our results to obtain is that medical expenditure is sensitive to costs.[6] We also abstract away from any costs associated with the provision of insurance.

[5] Contracts are not directly contingent on initial health levels h_0, but this would make little difference here as individuals are ex-ante identical.

[6] Exactly how spending responds to costs is a difficult empirical question that has been addressed in the literature. The response depends on a number of questions not discussed in this paper, including dynamic considerations. We refer the reader to Einav and Finkelstein (2018) for a discussion of many of these issues.

2.3 Health Care Providers

Health care providers maximize profits and can transform borrowed capital b' into medical services that are sold at the market price p, with production function $F(\cdot)$ such that $F'' < 0 < F'$.[7] There is a measure 1 of providers, and thus we ignore the possibility of entry or exit. Funds can also be invested in an alternative use with constant per unit returns $G > 1$. Denote the investment in medical care by I, and the investment in the alternative use by I'. Thus, producers face the following simple budget constraint $I + I' = b'$, while profits are

$$\Pi = pF(I) + GI' - Rb'. \tag{1}$$

As in Gaynor et al. (2000), we assume that individual providers are price-takers, which abstracts from the often convoluted and opaque process by which prices are formed in practice. Generally, the price of some medical service may be determined by government fiat, bargaining between health and insurance providers, the result of a more impersonal market process, or some combination of the like. This presents a challenge, especially since health care in the model can represent many distinct goods and services, which can be priced in a variety of ways depending on the good and the location. For some goods, the competition assumption may be a good approximation, while for others this is likely not the case. Some services are highly specialized and local, but even in this case if there is government regulation or an insurer with a powerful bargaining position, this would limit supplier market power. For example, the cost of some procedure may be determined by a bargaining process that begins with the average cost of that procedure across the industry within a country or a group of countries. In such a case, a model of pricing could be very complex and reflect dynamics, local market power, or cost considerations. However, it is reasonable to assume that market forces outside of the providers' control, i.e., aggregate supply and demand, would

[7] It is likely (we hope) that providers' objectives are not defined solely by the maximization of profit. Other motivations are not pertinent here as long as there is at least some profit motive such that medical care provision is responsive to the price.

still play a significant role. Providing a richer model of pricing for these types of health services is beyond the scope of this paper and represents an opportunity for further study.

We also note that unlike much of the previous literature such as Gaynor et al. (2000) and Wigger and Anlauf (2007), we assume that supply is upward sloping and that production is not characterized by constant marginal costs. In our model, a linear production function would pin down the price of care and thus precludes the type of inefficiency discussed below.

3 Equilibrium

We define an equilibrium as an insurance contract, allocation of consumption, health care services, and investments. At the equilibrium allocation, individuals maximize their utility in each period, insurance firms break-even on all contracts, health care providers maximize profits, and markets for consumption in both periods, insurance, capital, and medical care all clear. We now characterize the equilibrium allocation, beginning with the optimal behavior of health-care providers, then individuals, and finally insurers.

3.1 Health Care Providers

Providers choose their investments and thereby their production schedules to maximize profits. We assume an interior solution, such that some investment in both technologies is always worthwhile. The optimality conditions are as follows:

$$pF'(I) - G = 0 \tag{2}$$

$$G = R, \tag{3}$$

which pins down both the investment in medical services $I(p, G)$ and the return to capital $R = G$. Our assumptions ensure that the supply of health care is increasing in the price and decreasing in the cost of financing, as described in the following result.

LEMMA 1. *The optimal choice of investment in the production of medical care, denoted $\hat{I}$, is increasing in the price of care and decreasing in the return on the outside option:*

$$\frac{\partial \hat{I}(p,G)}{\partial p} > 0 > \frac{\partial \hat{I}(p,G)}{\partial G}. \tag{4}$$

3.2 Individual's Problem

Expected utility of individuals is given by

$$U = u(c_0, h_0) + \beta \left[(1-\pi) u(c_1^h, h_0) + \pi v(c_1^s, h_0 - S + m) \right], \tag{5}$$

where c_1^h and c_1^s denote consumption at time 1 when healthy and sick respectively, and $\beta \in (0,1)$ is a discount factor. Individual's maximize (5) subject to the following set of budget constraints:

$$c_0 = y_0 - b - q \tag{6}$$

$$c_1^h = y_1 + Rb + \Pi \tag{7}$$

$$c_1^s = y_1 - pm(1-\phi) + Rb + \Pi. \tag{8}$$

where Π are the profits of health care suppliers and also the individual share of these as we have assumed a unit measure of ex-ante identical individuals. It is straightforward to show that insurance is worthwhile at the equilibrium, and thus the optimal choices of saving and medical expenditure at an interior solution are characterized as follows:

$$b: \quad u_c(c_0, h_0) = \beta R \left[(1-\pi) u_c(c_1^h, y_0) + \pi v_c(c_1^s, h_0 - S + m) \right] \tag{9}$$

$$m: \quad v_c(c_1^s, h_0 - S + m) p (1-\phi) = v_h(c_1^s, h_0 - S + m). \tag{10}$$

Note that from (10), it is clear that the marginal rate of substitution will not be equal to the price ratio since the latter is distorted by the co-pay factor $(1-\phi)$. The welfare loss associated with this price wedge is the focus of much of the existing literature that was described above, which usually assumes that a welfare loss arises because those that are ill consider premium costs to be sunk when purchasing health care (e.g., Pauly (1968)). Furthermore, it is reasonable to assume that these welfare losses (felt through higher premiums) are imperceptible to consumers since losses are often subsidized by employers or taxpayers. It is important to note that this is not the focus of this paper, since we effectively assume that consumers can foresee the premium implications of any copay arrangement. What is driving the inefficiency described below is the failure of insurers to anticipate the impacts of future spending on the price of care, i.e., the general equilibrium implications of changes to aggregate demand from the insurance contract. In deriving our main result, we shall make use of the following result. An exposition of the individuals' problem, including sufficient conditions for Lemma 2, can be found in Appendix A.

LEMMA 2. *Optimal spending on care $m(\phi, p)$, is increasing in the subsidy, and decreasing in the price of care.*

3.3 Insurance Contracts

Denote the value function associated with the individual's problem by $V(\phi, q, p)$. A competitive insurance provider solves the following problem:

$$\max_{\phi,q} V(\phi, q, p) \tag{11}$$

subject to the non-negative profit condition:

$$q \geq \pi\phi pm, \tag{12}$$

where q represents the aggregate premiums and $\pi\phi pm$ is the aggregate expenditure of individuals that are ill. Substituting for the premium using (12) and differentiating with respect to ϕ yields the following optimal condition

$$u_c(c_0, h_0)\,(1 + \epsilon_\phi) = \beta v_c(c_1^s, h_0 - S + m), \tag{13}$$

where $\epsilon_\phi = \frac{\phi}{m}\frac{\partial m}{\partial \phi}$, is the elasticity of medical expenditure with respect to the subsidy. From (13) we see that the marginal utility of income for the sick is higher, the more responsive health care spending is to the subsidy, i.e., the larger is ϵ_ϕ.

4 Welfare Analysis

In this section we consider the efficiency of the competitive equilibrium described above. It is helpful to first describe the efficient allocation as a benchmark.

4.1 The efficient allocation

Consider a planner that maximizes ex-ante individual utility, subject to the relevant resource constraints. Denoting the planner's optimal choice variables with a hat, but otherwise keeping the notation as above, we formalize this problem as one of maximizing

$$U = u(c_0, h_0) + \beta\left[(1-\pi)u(c_1^h, h_0) + \pi v(c_1^s, h_0 - S + m)\right], \tag{14}$$

subject to the following set of resource constraints:

$$c_0 + b = y_0, \tag{15}$$

$$(1-\pi)c_1^h + \pi c_1^s = y_1 + G(b - I), \tag{16}$$

$$\pi m = F(I). \tag{17}$$

The efficient allocation is characterized by:

$$u_c(\hat{c_0}, h_0) = \beta G u_c(\hat{c_1}^h, h_0), \tag{18}$$

$$u_c(\hat{c_1}^h, h_0) = v_c(\hat{c_1}^s, h_0 - S + \hat{m}), \tag{19}$$

$$\frac{v_c(\hat{c_1}^s, h_0 - S + \hat{m})}{v_h(\hat{c_1}^s, h_0 - S + \hat{m})} = \frac{F'(\hat{I})}{G}. \tag{20}$$

We see that the marginal utility of income over time is proportional to the returns to saving and the discount factor, while the marginal rate of substitution between health care and consumption reflects the relevant costs. Contrasting these conditions with (9)-(10) and (13), the reader can observe that the competitive equilibrium described above is generally inefficient.

4.1.1 Indemnity Policy

To implement the efficient allocation described above, consider a simple insurance contract in which the insured are assumed to make a payment at $t = 0$, and receive the fixed payment M in the event of illness. This type of contract is often referred to as an indemnity. The zero expected profit condition on the insurance contract pins down the initial payment to $\pi M/R$. Analyzing the competitive equilibrium described above with this type of contract is very similar, and in particular (9)-(10) are the same except the price distortion caused by ϕ is absent. Further, (13) is also the same except that there is no elasticity term because the transfer is lump-sum. Combining these with the equilibrium behavior of the firm, and it is clear that the equilibrium allocation is first-best, as characterized above.

4.2 Efficiency at the competitive equilibrium

The indemnity contract described above is generally not feasible. As discussed above, the linear structure of health insurance contracts that tend to be observed in practice give rise to inefficiency since markets for risk are incomplete. In this section we consider the pricing implications of changes to our environment, and then the impact of perturbing the aggregate copay ϕ at the competitive allocation. Since the allocation is constrained inefficient, the welfare implications of these changes are not immediately obvious. The following result characterizes the effect of a change in the copay on the price of care.

PROPOSITION 1. *The price of care is increasing in ϕ.*

Proof. See Appendix B. □

Thus, a reduction in individual medical risk increases the cost of care not only through increased insurance premiums, but also by driving up the price of care. The main result is summarized in the following proposition.

PROPOSITION 2. *The competitive allocation defined in Section 3 is characterized by excessive insurance coverage, in that a reduction in the insurance subsidy ϕ is welfare improving.*

Proof. See Appendix B. □

The Proposition describes the welfare implications of perturbing the copay at the market allocation. The technical details are outlined in the appendix, but the intuition is fairly straightforward. Individual choices of medical care reflect the incentives embedded in the insurance contract. This drives a wedge between the value of care on the supply and demand side of the market. Furthermore, insurance contracts maximize the utility of consumers while taking the price of care as given, so not only does a subsidy impose an inefficiency at the margin, but the subsidy is not chosen with full consideration of price effects in a general equilibrium. As a result, the copay is excessive and welfare is improved by reducing coverage, and in turn the price of care.

5 Conclusion

A health insurance contract which provides lump-sum payments for all contingencies is efficient, but not generally observed in practice. Instead, we tend to observe variants of the contract described in Figure 3, a simplified version of which we take as given in this paper. Under relatively mild conditions, the contract structure implies the existence of a pecuniary externality in a general equilibrium setting. Thefore, the competitive equilibrium is generally not even second-best, and insurance contracts provide too much coverage at the margin. In fact, polices that reduce insurance coverage can increase welfare at the margin. This is particularly relevant when one considers the tremendous growth in health insurance and spending on care in recent years.

References

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein, 2015, Behavourial Moral Hazard in Insurance, *Quarterly Journal of Economics* 1623-1667, 877–921.

Chernew, Michael E., Allison B. Rosen, and A. Mark Fendrick, 2007, Value-Based Insurance Design, *Health Affairs* 26, 195–203.

Cutler, David, and Richard Zeckhauser, 2000, *The Anatomy of Health Insurance*, volume 1A of *Handbook of Health Economics*, 563–643 (Elsevier).

Einav, Liran, and Amy Finkelstein, 2018, Moral Hazard in Health Insurance: What We Know and How We Know it, *Journal of the European Economic Association* 16 (4), 957–982.

Feldman, Roger, and Bryan Dowd, 1991, A new estimate of the welfare loss of excess health insurance, *American Economic Review* 81, 297–301.

Feldstein, Martin, 1973, The welfare loss of excess health insurance, *Journal of Political Economy* 81, 251–80.

Feldstein, Martin, and Bernard Friedman, 1977, Tax subsidies, the rational demand for insurance and the health care crisis, *Journal of Public Economics* 7, 155–178.

Finkelstein, Amy, 2007, The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare, *Quarterly Journal of Economics* 122, 1–37.

Gaynor, Martin, Deborah Haas-Wilson, and William B. Vogt, 2000, Are Invisible Hands Good Hands? Moral Hazard, Competition, and the Second-Best in Health Care Markets, *Journal of Political Economy* 108, 992–1005.

Geanakoplos, John, and Herakles Polemarchakis, 1986, Existence, regularity, and constrained suboptimality of competitive allocations when the asset market is incomplete, in W.P. Heller, R.M. Ross, and D.A. Starrett, eds., *Uncertainty, Information and Communication: Essays in Honor of Kenneth Arrow*, volume 3, 65–95 (Cambridge University Press).

Greenwald, Bruce, and Joseph Stiglitz, 1986, Externalities in Economies with Imperfect Information and Incomplete Markets, *Quarterly Journal of Economics* 101, 229–264.

Gruber, J., and H. Levy, 2009, The Evolution of Medical Spending Risk, *Journal of Economic Perspectives* 23, 25–48.

Hart, Oliver, 1975, On the Optimality of Equilibrium when the Market Structure is Incomplete, *Journal of Economic Theory* 11, 418–443.

Pauly, Mark V., 1968, The Economics of Moral Hazard: Comment, *American Economic Review* 58, 531–537.

Pauly, Mark V., and F.E. Blavin, 2008, Moral hazard in insurance, value-based cost sharing, and the benefits of blissful ignorance, *Journal of Health Economics* 27, 1407–1417.

Wigger, Berthold, and Markus Anlauf, 2007, Do consumers purchase too much health insurance? The role of market power in health-care markets, *Journal of Public Economic Theory* 9, 547–561.

Zweifel, Peter, and Willard Manning, 2000, *Moral Hazard and Consumer Incentives in Health Care*, volume 1A of *Handbook of Health Economics*, 409–459 (Elsevier).

Appendix A The Individual's Problem

For convenience, the individuals objective function, along with the first order conditions associated with an interior solution are rewritten below.

$$U(b,m) = u(c_0, h_0) + \beta \left[(1-\pi) u(c_1^h, h_0) + \pi v(c_1^s, h_1^s) \right], \tag{21}$$

where

$$c_0 = y_0 - b - q, \tag{22}$$

$$c_1^h = y_1 + Rb + \Pi, \tag{23}$$

$$c_1^s = y_1 - p(1-\phi)m + Rb + \Pi, \tag{24}$$

$$h_1^s = h_0 - S + m. \tag{25}$$

The first-order conditions are:

$$b: \quad u_c(c_0, h_0) = \beta R \left[(1-\pi) u_c(c_1^h, h_0) + \pi v_c(c_1^s, h_1^s) \right],$$

$$m: \quad v_c(c_1^s, h_1^s) p (1-\phi) = v_h(c_1^s, h_1^s).$$

To reduce clutter, define

$$U_{bb} = u_{cc}(c_0, h_0) + \beta R^2 \left[(1-\pi) u_{cc}(c_1^h, h_0) + \pi v_{cc}(c_1^s, h_1^s) \right], \tag{26}$$

$$U_{mm} = \beta \pi \left[p^2 (1-\phi)^2 v_{cc} - 2p(1-\phi) v_{hc} + v_{hh} \right], \tag{27}$$

$$U_{bm} = U_{mb} = \beta R \pi \left[-p(1-\phi) v_{cc} + v_{ch} \right]. \tag{28}$$

Given our assumptions on utility, negative definiteness implies that

$$U_{bb} < 0 \text{ and } U_{bb} U_{mm} - U_{bm} U_{mb} > 0 \implies U_{mm} < 0. \tag{29}$$

A.1 Signing $\frac{\partial m}{\partial \phi}$

Differentiating the first-order conditions of the individuals' problem with respect to ϕ we obtain:

$$u_{cc}(c_0, h_0)\frac{\partial c_0}{\partial \phi} = \beta R\left[(1-\pi)u_{cc}(c_1^h, h_0)\frac{\partial c_1^h}{\partial \phi} + \pi v_{cc}(c_1^s, h_1^s)\frac{\partial c_1^s}{\partial \phi} + \pi v_{ch}(c_1^s, h_1^s)\frac{\partial m}{\partial \phi}\right],$$
$$\left[v_{cc}(c_1^s, h_1^s)\frac{\partial c_1^s}{\partial \phi} + v_{ch}(c_1^s, h_1^s)\frac{\partial m}{\partial \phi}\right]p(1-\phi) - pv_c(c_1^s, h_1^s) =$$
$$\left[v_{hc}(c_1^s, h_1^s)\frac{\partial c_1^s}{\partial \phi} + v_{hh}(c_1^s, h_1^s)\frac{\partial m}{\partial \phi}\right],$$

where

$$\frac{\partial c_0}{\partial \phi} = -\frac{\partial b}{\partial \phi},\ \frac{\partial c_1^h}{\partial \phi} = R\frac{\partial b}{\partial \phi},$$
$$\frac{\partial c_1^s}{\partial \phi} = -p(1-\phi)\frac{\partial m}{\partial \phi} + pm + R\frac{\partial b}{\partial \phi},\ \frac{\partial h_1^s}{\partial \phi} = \frac{\partial m}{\partial \phi}.$$

Using the first equation we obtain:

$$\frac{\partial b}{\partial \phi} = \frac{-\beta R\pi\left[-p(1-\phi)v_{cc} + v_{ch}\right]\frac{\partial m}{\partial \phi} - \beta R\pi p m v_{cc}}{u_{cc}(c_0, h_0) + \beta R^2\left[(1-\pi)u_{cc}(c_1^h, h_0) + \pi v_{cc}(c_1^s, h_1^s)\right]}, \tag{30}$$

$$= \frac{-U_{bm}\frac{\partial m}{\partial \phi} - \beta R\pi p m v_{cc}}{U_{bb}}. \tag{31}$$

Using this in the second we have:

$$\frac{\partial m}{\partial \phi} = \frac{pv_c + (-p(1-\phi)v_{cc} + v_{hc})\left(pm + R\frac{\partial b}{\partial \phi}\right)}{p^2(1-\phi)^2 v_{cc} - 2p(1-\phi)v_{hc} + v_{hh}}, \tag{32}$$

$$= \frac{-\beta\pi pv_c - \beta\pi\left(-p(1-\phi)v_{cc} + v_{hc}\right)\left(pm + R\frac{\partial b}{\partial \phi}\right)}{U_{mm}}, \tag{33}$$

$$= \frac{-v_c p\beta\pi U_{bb} - U_{bm}U_{bb}pm/R + U_{bm}v_{cc}\beta R\pi pm}{U_{bb}U_{mm} - U_{bm}U_{mb}}, \tag{34}$$

$$= \frac{-v_c p\beta\pi U_{bb} - \frac{pmU_{bm}}{R}\left(U_{bb} - v_{cc}\beta R^2\pi\right)}{U_{bb}U_{mm} - U_{bm}U_{mb}}, \tag{35}$$

$$= \frac{-v_c p\beta\pi U_{bb} - \frac{pmU_{bm}}{R}\left(u_{cc}(c_0, h_0) + \beta R^2(1-\pi)u_{cc}(c_1^h, h_0)\right)}{U_{bb}U_{mm} - U_{bm}U_{mb}} > 0. \tag{36}$$

A sufficient condition for the final inequality is thus $U_{bm} \geq 0$. See discussion below.

A.2 Signing $\frac{\partial m}{\partial p}$

Differentiating the first-order conditions with respect to p we obtain:

$$u_{cc}(c_0, h_0)\frac{\partial c_0}{\partial p} = \beta R\left[(1-\pi)u_{cc}(c_1^h, h_0)\frac{\partial c_1^h}{\partial p} + \pi v_{cc}(c_1^s, h_1^s)\frac{\partial c_1^s}{\partial p} + \pi v_{ch}(c_1^s, h_1^s)\frac{\partial m}{\partial p}\right], \tag{37}$$

$$\left[v_{cc}(c_1^s, h_1^s)\frac{\partial c_1^s}{\partial p} + v_{ch}(c_1^s, h_1^s)\frac{\partial m}{\partial p}\right]p(1-\phi) + (1-\phi)v_c(c_1^s, h_1^s) = \left[v_{hc}(c_1^s, h_1^s)\frac{\partial c_1^s}{\partial p} + v_{hh}(c_1^s, h_1^s)\frac{\partial m}{\partial p}\right], \tag{38}$$

where

$$\frac{\partial c_0}{\partial p} = -\frac{\partial b}{\partial p},\ \frac{\partial c_1^h}{\partial p} = R\frac{\partial b}{\partial p}, \tag{39}$$

$$\frac{\partial c_1^s}{\partial p} = -(1-\phi)m - p(1-\phi)\frac{\partial m}{\partial p} + R\frac{\partial b}{\partial p},\ \frac{\partial h_1^s}{\partial p} = \frac{\partial m}{\partial p}. \tag{40}$$

Using the first equation we obtain:

$$\frac{\partial b}{\partial p} = \frac{-\beta R\pi\left[-p(1-\phi)v_{cc}+v_{ch}\right]\frac{\partial m}{\partial \phi} - \beta R\pi p m v_{cc}}{u_{cc}(c_0,h_0)+\beta R^2\left[(1-\pi)u_{cc}(c_1^h,h_0)+\pi v_{cc}(c_1^s,h_1^s)\right]}, \tag{41}$$

$$= \frac{-U_{bm}\frac{\partial m}{\partial p}+\beta R\pi(1-\phi)m v_{cc}}{U_{bb}} \tag{42}$$

Using this in the second we have:

$$\frac{\partial m}{\partial p} = \frac{v_c(1-\phi)\beta\pi U_{bb} + \frac{U_{bm}U_{bb}(1-\phi)m}{R} - U_{bm}v_{cc}(1-\phi)m\pi\beta R}{U_{bb}U_{mm}-U_{bm}U_{mb}}, \tag{43}$$

$$= \frac{v_c(1-\phi)\beta\pi U_{bb} + U_{bm}(1-\phi)m\left(\frac{u_{cc}(c_0,h_0)}{R}+\beta R(1-\pi)u_{cc}(c_1^h,h_0)\right)}{U_{bb}U_{mm}-U_{bm}U_{mb}} < 0 \text{ if } U_{bm}\geq 0. \tag{44}$$

Since $U_{bm} = \beta R\pi\left[-p(1-\phi)v_{cc}+v_{ch}\right]$, we have the sufficient condition $v_{ch} \geq 0$, which is an unnecessarily strong assumption that we do not make in the paper since this is quite tangential and we do not wish to weigh in on the law of demand here. Note that comparing (36) and (43), we can see that $\frac{\partial m}{\partial p} = -\frac{\partial m}{\partial \phi}\frac{(1-\phi)}{p}$ as might be expected since the effective price is $\tilde{p} = p(1-\phi)$.

Appendix B Proofs

B.1 Proof of Proposition 1

Consider the market clearing condition $F(I^*) = \pi m$. Differentiating with respect to ϕ we obtain:

$$F'(I^*)\frac{\partial I^*}{\partial p}\frac{dp}{d\phi} = \pi\left[\frac{\partial m^*}{\partial p}\frac{dp}{d\phi} + \frac{\partial m^*}{\partial \phi}\right]. \tag{45}$$

Therefore,

$$\frac{dp}{d\phi} = \frac{\frac{\partial m^*}{\partial \phi}}{\frac{F'(I^*)}{\pi}\frac{\partial I^*}{\partial p} - \frac{\partial m^*}{\partial p}} > 0, \tag{46}$$

since $\frac{\partial I^*}{\partial p} > 0$, $\frac{\partial m^*}{\partial \phi} > 0$ and $\frac{\partial m^*}{\partial p} < 0$ from Lemmas 1 and 2.

B.2 Proof of Proposition 2

Differentiating the ex-ante utility of households with respect to ϕ, we have

$$\begin{aligned}\frac{dU}{d\phi} = -u_c(c_0,h_0)\frac{\partial q}{\partial p}\frac{dp}{d\phi} + \beta\Bigg[&(1-\pi)u_c(c_1^h,y_0)\frac{\partial \Pi}{\partial p}\frac{dp}{d\phi} + \pi v_c(c_1^s,\cdot)\frac{\partial \Pi}{\partial p}\frac{dp}{d\phi} + \\ &\pi v_c(c_1^s,\cdot)\left(-m(1-\phi)\frac{dp}{d\phi} - p(1-\phi)\frac{\partial m}{\partial p}\frac{dp}{d\phi}\right) + \pi v_h(c_1^s,\cdot)\frac{\partial m}{\partial p}\frac{dp}{d\phi}\Bigg].\end{aligned} \tag{47}$$

Using the first-order conditions for the household problem, we can write the above expression as follows:

$$\frac{dU}{d\phi} = -u_c(c_0,h_0)\frac{\partial q}{\partial p}\frac{dp}{d\phi} + \frac{1}{R}u_c(c_0,y_0)\frac{\partial \Pi}{\partial p}\frac{dp}{d\phi} - \beta\pi m(1-\phi)v_c(c_1^s,\cdot)\frac{dp}{d\phi}. \tag{48}$$

Applying the optimality condition for the insurer and factoring out the price effect yields

$$\frac{dU}{d\phi} = \left[-\frac{\partial q}{\partial p} + \frac{1}{R}\frac{\partial \Pi}{\partial p} - \pi m(1-\phi)\left(1+\epsilon_\phi\right)\right]u_c(c_0,h_0)\frac{dp}{d\phi}. \tag{49}$$

We replace the partial derivatives with respect to p using

$$\frac{\partial q}{\partial p} = \pi\phi m + \pi\phi p \frac{\partial m}{\partial p} \tag{50}$$

$$\frac{\partial \Pi}{\partial p} = F(I^*) = \pi m, \tag{51}$$

where the last equality follows from equating aggregate supply of medical care with aggregate demand. Using these in the expression for the perturbation effect above we obtain:

$$\frac{dU}{d\phi} = \left[-\pi\phi m - \pi\phi p \frac{\partial m}{\partial p} + \frac{1}{R}\pi m - \pi m(1-\phi)\left(1+\epsilon_\phi\right)\right] u_c(c_0, h_0)\frac{dp}{d\phi}, \tag{52}$$

$$= \pi m \left[-\left(\phi + (1-\phi)\left(1+\epsilon_\phi\right)\right) + \left(\frac{1}{R} - \phi\epsilon_p\right)\right] u_c(c_0, h_0)\frac{dp}{d\phi}, \tag{53}$$

$$= \pi m \left[\left(\frac{1}{R} - 1\right) - \left(\phi\epsilon_p + (1-\phi)\epsilon_\phi\right)\right] u_c(c_0, h_0)\frac{dp}{d\phi}, \tag{54}$$

$$= \pi m \left(\frac{1}{R} - 1\right) u_c(c_0, h_0)\frac{dp}{d\phi}, \tag{55}$$

as it can be shown that $\phi\epsilon_p + (1-\phi)\epsilon_\phi = 0$. The above expression clearly shows that perturbation effect must have the opposite sign as the price effect since $R > 1$. The price effect is shown to be positive in Proposition 1.