# Health Insurance as a Two-Part Pricing Contract [*]

**Darius Lakdawalla and Neeraj Sood**

**RAND Corporation and NBER**

November 20, 2008

Abstract

Monopolies appear throughout health care, as a result of patents, limits to the extent of the market, or the presence of unique inputs and skills. However, the presence of health insurance alters the welfare implications and analysis of monopoly. Health insurance, often implemented as an ex ante premium coupled with an ex post co-payment per unit consumed, effectively operates as a two-part pricing contract that allows monopolists to extract profits from consumers without inefficiently constraining quantity. Frictionless and competitive health insurance markets, even under incomplete or asymmetric information, perfectly eliminate deadweight losses from monopoly in health care. Frictions or regulations that limit the ability to "premium-discriminate" may restore some of these monopoly losses, which will then manifest as uninsurance for the healthy (or poor), or under-insurance for the chronically ill (or rich). Finally, health insurance may also improve dynamic incentives to innovate. Empirical analysis of pharmaceutical patent expiration confirms that heavily insured markets experience little to no gain in welfare when market power falls, and that health insurance helps patent monopolists extract more surplus than they could in less insured markets.

---

[*] For helpful comments and suggestions, we wish to thank Alan Garber, Dana Goldman, Geoffrey Joyce, Tomas Philipson, Bill Vogt, participants in the 2005 NBER Summer Institute Health Economics Workshop, participants in the Fall 2006 NBER Health Care Program Meetings, the University of Illinois, and the University of Chicago. We are grateful to the National Institute on Aging for funding.

## A. Introduction

Optimal health insurance contracts balance risk-sharing against the need for efficient utilization incentives (Arrow, 1963; Pauly, 1968; Zeckhauser, 1970). This balance explains why such contracts do not entitle policyholders to unlimited utilization, but instead charge an ex post unit price or co-payment. Co-payments reduce the degree of insurance, but in return produce fewer distortions in the goods market, because the consumer faces a private price that partially reflects social cost.

The trade-off between risk-sharing and incentives has been widely studied. However, health insurance contracts have another function that is less well-appreciated: the reduction of deadweight loss. Health insurance resembles a two-part pricing contract, in which a group of consumers pays an upfront fee in exchange for lower prices in the event of illness. Two-part pricing contracts allow a monopolist to sell goods at marginal cost, but extract consumer surplus in the form of an upfront payment (see the seminal paper by Oi, 1971). Typically, competition improves consumer welfare, because it minimizes deadweight loss. In the presence of two-part pricing, however, the monopolist has the same incentive to minimize deadweight loss, because such a strategy maximizes the consumer surplus available for extraction.

An example illustrates this mechanism in the insurance context. Setting a marginal cost co-payment and a premium equal to expected consumer surplus would, for instance, allow a firm to extract the maximum possible surplus, while still ensuring efficient utilization and zero deadweight loss. The uncertainty of health care demand, coupled with ex ante or ex post asymmetric information, *creates* a contractual structure that facilitates more efficient extraction of consumer surplus.

We show that, as a general matter, frictionless and competitive health insurance markets will perfectly eliminate deadweight losses from market power in health care provision. This logic is robust to typical informational failures like moral hazard and adverse selection. Deadweight losses from monopoly arise only when regulations or other frictions create conditions under which insurers – by choice or mandate – offer pooled and uniform contracts to heterogeneous consumers.

Our results have several important implications for evaluating monopoly loss in health care markets. First, the extent and even presence of deadweight losses from health care monopoly are determined by the structure of the insurance market. Therefore, the decision to regulate or allow monopoly in health care must be driven in large part by the structure of the health insurance market that mediates the provision of goods. Moreover, extending insurance coverage or promoting efficiency in the insurance market may be viewed as substitutes for regulating monopoly in health care markets.

Second, the price-cost margin is a less reliable measure of welfare loss in markets where health insurance plays a role. When insurance is widespread and reasonably complete, providers may be receiving very high profits and prices, even though consumers are paying prices near marginal cost. The only reliable measure of market power is the change in quantity it induces, or, possibly, the change in rates of uninsurance or under-insurance.

Finally, health insurance lowers the static deadweight losses associated with patent monopolies. Health care markets may thus not be as subject to the usual societal trade-off between incentives to innovate and static inefficiency.

Empirical analysis of patent expiration in the pharmaceutical market provides evidence consistent with these results. First, the elimination of patent monopoly has little to no impact on quantity consumed for molecules that are heavily insured, but substantial impact for molecules with less widespread insurance. Second, even though market quantity barely falls for heavily insured molecules, the profits or revenues earned by the innovating firm falls by more than in less insured markets. This is consistent with the notion that health care monopolists implicitly use health insurance as a means of extracting more consumer surplus.

Section B develops our argument that competitive and frictionless insurance markets eliminate deadweight loss from monopoly, even when information is incomplete, and market power is imperfect. Section C explores how contracting frictions and other incentives for insurance pooling ultimately drive deadweight losses from health care monopoly. Section D analyzes the implications for innovation and patent protection. Section E presents our empirical analysis. Finally, Section E summarizes our conclusions and implications for the analysis of market power in health care.

## B. Two-Part Health Insurance and Surplus-Extraction

Any insurer who can charge both a premium ex ante and a co-payment ex post has enough tools to extract maximum consumer surplus and ensure efficient utilization. We first make this point in the context of a standard model with moral hazard. Our initial setup is very similar to that of Gaynor, Haas-Wilson, and Vogt (2000), who show that reductions in the price of medical care benefit consumers even in the presence of moral hazard.

Consumers face a risk of illness, and an uncertain demand for a medical remedy, produced at constant marginal cost $MC$. An insurance contract is an offer of an ex post co-payment ($m$), and a premium $I$. There are consumers of measure one, indexed by $h \in [0,1]$, and distributed uniformly over this interval. Patients with lower values of $h$ are sicker. The fraction $\sigma$ of consumers fall sick. Sick consumers place value on the medical remedy, while healthy consumers do not.

Consumers do not know their value of $h$ ex ante, but learn it ex post. Insurers, however, cannot observe this value and thus cannot make indemnity payments conditional on the underlying health state. Payments can only be contingent on the consumer's observed decision to purchase the medical good or not. The necessity of tying payments to utilization, rather than the underlying source of risk, results in "over-utilization" relative to the first-best, full information case. This is a second-best means of delivering some additional insurance in the face of informational incompleteness.

## B.1 The Typical Competitive Problem

We begin by characterizing the standard competitive equilibrium allocation in the presence of moral hazard. Consider a representative competitive insurer purchasing medical care from a competitive goods market selling at marginal cost, and providing insurance within the informational structure outlined above. The firm chooses a co-payment and premium that maximizes consumer utility, subject to a break-even constraint, and incentive compatibility for the consumer. The insurer knows the quantity of medical care demanded by consumer $h$, given the co-payment and income, according to $q(W-I,m,h)$.

The firm's optimization problem can be written as:

$$\begin{aligned}&\max_{I,m\leq MC}\int_0^1 u(W-I-mq^*,q^*,h)dh\\&s.t.\ I+(m-MC)E(q^*)\geq 0\\&and\ q^*=q(W-I,m,h)\end{aligned}\tag{1}$$

Associating the multiplier $\mu$ with the break-even constraint, the first-order conditions can be expressed as:

$$\begin{aligned}&[I]: Eu_W=\mu\big(1-(m-MC)E(q_W)\big)\\&[m]:\int_0^1\frac{q}{E(q)}u_W(W-I-mq,q,h)dh=\mu\left(1-(m-MC)\frac{-E(q_m)}{E(q)}\right)\end{aligned}\tag{2}$$

These first-order conditions illustrate the standard trade-off between risk-bearing and incentives in the presence of moral hazard. The left-hand side of the first order condition for $m$ always exceeds the left-hand side of the condition for $I$, because $u_W$ and $q^*$ are decreasing in $h$.[1] This fact, coupled with the two first-order conditions, implies that

---

[1] $\int_0^1\frac{q}{E(q)}u_W(W-I-mq,q,h)dh$ is a weighted average of $u_W$, where more weight is placed on its larger values.

$$(m - MC)\left(\frac{E(q_m)}{E(q)} + E(q_W)\right) > 0 \quad (3)$$

Observe that $E(q_m) + E(q)E(q_W)$ is the expected effect on $q$ of a compensated increase in the co-payment $m$. Since the compensated demand for medical care is downward-sloping,[2] it follows that $\left(\frac{E(q_m)}{E(q)} + E(q_W)\right) < 0$, and $m < MC$. In turn, this implies that the marginal utility of wealth will be higher than in the first-best, according to the first-order condition for insurance.

Intuitively, when ex post information is asymmetric, the only way to provide insurance is to induce over-utilization by charging the consumer a price below marginal cost. Therefore, the benefits of insurance must be traded off against the cost of inducing distortion in the goods market. This leads to: (1) Over-utilization relative to the first-best, (2) Higher marginal utility of wealth relative to first-best, and (3) Incomplete insurance.

### B.2 Two-Part Health Insurance with Monopoly

Two-part health insurance eliminates deadweight losses associated with monopoly, but it cannot solve the intrinsic informational problems that lead to moral hazard in this environment. As a result, a monopolist with access to two-part health insurance pricing will choose an allocation of resources identical to the competitive benchmark described above. For initial simplicity, we consider an insurer who is also the monopoly provider of the good with uncertain demand. Later, we separate the two functions. Moreover, while we consider the simpler case of pure uncontested monopoly in the text, Appendix A generalizes our results to the case of incomplete market power.

---

[2] The consumer's first-order condition for medical care consumption is $E(-mu_W + u_q) = 0$. The first-order effect of a compensated increase in the co-payment is an increase in $m$, but no change in $u_W$ or $u_q$. Therefore, the consumer must decrease medical care consumption in response.

The insurer-provider maximizes profits subject to a reservation utility condition for the consumer. Define $\overline{U}$ as the level of utility the consumer would attain if he refused the insurance contract and failed to consume the medical care good. However, he may still have a claim on the firm's profits if he is a shareholder. The firm thus solves:

$$\begin{aligned} &\max_{I,m} I + (m - MC)E(q(W - I, m, h)) \\ &s.t. \int_0^1 u(W - I - mq + \pi, q, h)dh \geq \overline{U} \end{aligned} \tag{4}$$

It is straightforward to prove the mechanical equivalence between this problem and the competitive one. If we define $\overline{\pi}$ as the equilibrium monopoly profit level, the above problem is equivalent to:

$$\begin{aligned} &\max_{I,m} \int u(W - I - mq + \overline{\pi}, q, h)dh \\ &s.t. I + (m - MC)E(q) \geq \overline{\pi} \end{aligned} \tag{5}$$

By displacement, this can be rewritten as:

$$\max_m \int_0^1 u(W - m(q - E(q)) - MC * E(q), q, h)dh, \tag{6}$$

which is exactly equivalent to the displaced version of the problem in 1. Intuitively, the monopolist can best maximize his own profits by first maximizing gross consumer surplus, and then extracting it in the form of an upfront payment.

### B.3 Separating the Insurance- and Goods-Producers

Next, we separate insurers from goods-producers, and explore how our results generalize. Consider a representative insurer standing in for a competitive industry. The insurer faces a health care monopolist, who is able to specify both a price and a quantity, or equivalently, a quantity and a total fixed fee. This type of contracting is often observed in health care markets, where quantities are either explicitly named (e.g., by a pharmaceutical wholesaler), tied to a nonlinear price schedule (e.g., in the form of quantity discounts, rebates, and the like), or where prices are separately negotiated with providers of different sizes.

To offer a few examples, contracts between PBMs and pharmaceutical firms are of two types – non-capitated and capitated.[3] Non-capitated contracts usually specify a list price or "wholesale acquisition cost" and terms for determining discounts or rebates. Rebates are usually tied to the dollar or unit sales of a particular drug product. Growth rebates offer PBMs a steeper discount if they achieve certain volume targets. Capitated contracts, on the other hand, specify a fixed payment from the PBM to the drug company per insured member per month, along with some risk-sharing arrangement that determines additional payments or concessions based on actual drug usage (Levy, 1999). The capitated rates combined with risk-sharing arrangements effectively render these equivalent to two-part pricing contracts. Outside the pharmaceutical industry, health insurers routinely negotiate different prices with different hospitals and providers, even for similar products and services. These negotiated prices depend on provider size, insurer size, the form of the insurer, and other market characteristics (Zwanziger and Melnick, 1988; Dranove et al., 1993; Keeler et al., 1999; Dor et al., 2004).

The ability to tie different prices to different quantities is important. When the monopolist is able to specify only one of these, we revert to the analysis of monopoly articulated by Gaynor, Haas-Wilson, and Vogt (2000), where the usual societal losses are incurred.[4] In the absence of two-part health insurance, specifying both prices and quantities for heterogeneous consumers is quite impractical. The provider would need to specify a different price-quantity pair, or two-part pricing

[3] Private-sector entities that offer prescription drug insurance coverage, such as employers, labor unions, and managed care companies, often hire pharmacy benefit managers (PBMs) to manage these insurance benefits. PBMs engage in many activities to manage their clients' prescription drug insurance coverage including assembling a network of retail pharmacies, designing the plan formulary and cost sharing arrangements (co-payments for different drugs) and negotiating with pharmaceutical companies.

[4] They show that even in the presence of moral hazard, consumers are better off with competition (lower prices) than with monopoly (higher prices).

menu, for each consumer. The two-part structure of health insurance provides a natural and practical way to tie price and quantity together.

The monopolist provider offers an average quantity $EQ$ and fixed-fee $F$, which solve the following profit-maximization problem:

$$\begin{aligned} &\max_{EQ,F} F - MC * EQ \\ &s.t. F \leq G(EQ) \end{aligned} \tag{7}$$

The monopolist cannot extract more than $G(EQ)$, defined as the maximum amount of expected gross profit the insurer is able to earn on the expected quantity $EQ$.

Gross profits can be characterized as the maximum expected amount the insurer can extract from consumers at that expected quantity level, or:

$$\begin{aligned} &G(EQ) = \max_{I,m} I + mE(q) \\ &s.t. \int_0^1 u(W - I - mq(m, W - I, h) + \pi, q(m, W - I, h), h)dh \geq \overline{U} \\ &and\ E(q) \leq EQ \end{aligned} \tag{8}$$

Defining $\eta$ as the Lagrange multiplier on the second constraint, the insurer's problem can be rewritten as:

$$\begin{aligned} &G(EQ) = \max_{I,m} I + mE(q) + \eta(EQ - E(q)) \\ &s.t. \int_0^1 u(W - I - mq(m, W - I, h) + \pi, q(m, W - I, h), h)dh \geq \overline{U} \end{aligned} \tag{9}$$

According to the envelope theorem, $G'(EQ) = \eta$ in this problem. Moreover, the first-order condition for the monopolist implies that $G'(EQ) = MC$. Since this implies $\eta = MC$ in equilibrium, we can write:

$$\begin{aligned} &G(EQ) = \max_{I,m} I + (m - MC)E(q) + MC * EQ \\ &s.t. \int_0^1 u(W - I - mq(m, W - I, h) + \pi, q(m, W - I, h), h)dh \geq \overline{U} \end{aligned} \tag{10}$$

The equilibrium values of $I$ and $m$ are unaffected by the constant term $MC * EQ$. Therefore, the problem above is equivalent the integrated insurer's problem in equation 4. This demonstrates the result that monopolists who can tie quantity to a fixed payment will induce insurers to behave as if they were goods-monopolists themselves.

### B.4 Adverse Selection

The basic logic of health insurance as two-part pricing also holds up under another common failing of insurance markets — adverse selection.[5] Our analysis of moral hazard demonstrated that a monopolist with access to two-part health insurance can replicate the competitive equilibrium with moral hazard, or incomplete *ex post* information. Adverse selection adds incomplete *ex ante* information. In this case, the insurer can observe neither the severity of illness ex post, nor the ex ante differences in the propensity of consumers to fall ill.

As in the case of moral hazard, two-part pricing cannot remove the deadweight loss associated with asymmetric information, but it does remove all the incremental deadweight loss associated with monopoly. In other words, a monopolist with access to the two-part contract will do just as well as a competitive market, in the face of asymmetric information.

We assume there are chronically ill patients (type $C$), and not chronically ill patients (type $N$). Firms cannot observe consumer types. Define $\mu^C(h)$ and $\mu^N(h)$ as the measures of chronically ill and not chronically ill people across the interval $h \in [0,1]$. The distribution for the chronically ill is assumed to dominate the other in the first-order stochastic sense. An insurance contract is an ex ante insurance premium $(I)$, coupled with an ex post copayment $(m)$. The appendix demonstrates that, under these circumstances, competition is Pareto-equivalent to monopoly, when two-part health insurance contracts are used.

---

[5] This analysis is related to Stiglitz's (1977) finding that, in the presence of adverse selection, monopoly within insurance markets can dominate competition.

## C. Pooling Equilibria

The results for adverse selection rely on the canonical Rothschild-Stiglitz separating equilibrium response to asymmetric information. Indeed, Rothschild and Stiglitz proved the now widely known result that pooling equilibria are not generally possible in a frictionless, competitive, and unregulated insurance market. However, frictions in the labor market (cf, Crocker and Moran, 2003; Bhattacharya and Vogt, 2006; Fang and Gavazza, 2007), market power among insurers (Stiglitz, 1977), community-rating laws (cf, Adams, 2007), or restrictions on premium-differentiation (cf, Bhattacharya and Bundorf, 2005), can generate distortions that lead to pooling equilibria. In a pooling equilibrium, the premium is a less efficient instrument for surplus-extraction. This can lead to inefficiency and deadweight loss from market power in the provision of medical care, but not of the typical form.

Continue with our two types of consumers — $C$ and $N$ -- the chronically ill and not chronically ill. As before, the $C$ types are sicker and derive more consumer surplus from any given health insurance offer. However, through some combination of legal, informational, or labor market constraints, firm either choose or are compelled to offer a single premium and copayment schedule. Insurers cannot distinguish between the two types ex ante (adverse selection), nor can they distinguish ex post between patients with different levels of illness severity (moral hazard).

### C.1 Competitive Outcomes

We first characterize the competitive pooling equilibrium that serves as a benchmark. Suppose there is some set of characteristics that enables a pooling equilibrium -- labor market frictions, or regulation – where no other distortions exist. Insurance and medical care are both provided by competitive firms. The representative medical care provider sells goods at marginal cost. All consumers have reservation utility levels $\overline{U}$. For completeness, we also endow the two types with $s^C$ and $s^N$ shares of the insurer's profits.

The representative insurer maximizes:

$$\begin{aligned}
&\max_{I,m} \int_0^1 u^N(W-I-mq^N, q^N, h)\mu^N(h)dh \\
&s.t.\ 2I + (m-MC)E(q^C(W-I,m,h)) + (m-MC)E(q^N(W-I,m,h)) \geq 0 \\
&\int_0^1 u^N(W-I-mq^N, q^N, h)\mu^N(h)dh \geq \overline{U} \\
&\int_0^1 u^C(W-I-mq^C, q^C, h)\mu^C(h)dh \geq \overline{U}
\end{aligned}$$

If the reservation utility constraint for the $C$ types binds, the $N$ types will exit the market. Since we assume a pooling equilibrium possible by construction, this cannot be an equilibrium generated by competitive firms, who could make both types better off by insuring everyone. Therefore, the participation constraint for $C$ types will fail to bind, and the constraint for the $N$ types is analytically redundant. The benchmark equilibrium solves:

$$\begin{aligned}
&\max_{I,m} \int_0^1 u^C(W-I-mq^C, q^C, h)\mu^C(h)dh \\
&s.t.\ 2I + (m-MC)E(q^C(W-I,m,h)) + (m-MC)E(q^N(W-I,m,h)) \geq 0
\end{aligned} \tag{11}$$

### C.2 Effects of Market Power on Uninsurance

Now consider an integrated monopolist that provides insurance and medical care. The integrated firm offers a single contract to both types. Each type is endowed with $s^H$ and $s^L$ shares of the firm, which solves the following problem:

$$\begin{aligned}
&\max_{I,m} 2I + (m-MC)E(q^C(W-I,m,h)) + (m-MC)E(q^N(W-I,m,h)) \\
&s.t.\ \int_0^1 u^C(W-I-mq^C + s^C\pi, q^C, h)\mu^C(h)dh \geq \overline{U} \\
&\int_0^1 u^N(W-I-mq^N + s^N\pi, q^N, h)\mu^N(h)dh \geq \overline{U}
\end{aligned}$$

Define $(I,m)$ as the solution to this problem. Since there is a single contract, the firm will not be able to capture consumer surplus from both these consumer types, and one of the participation constraints will fail to bind. Moreover, if the equilibrium contract extracts all available surplus from the $C$ types, it will violate the participation constraint for the $N$ types. Therefore, there are two cases to consider: (1) the participation constraint binds for the $N$ types, and both types receive insurance; and (2) the participation constraint binds for the $C$ types, and only they receive insurance.

Begin with the first case, where both types are insured. The problem simplifies to:

$$\begin{aligned}&\max_{I,m} 2I+(m-MC)E(q^C(W-I,m,h))+(m-MC)E(q^N(W-I,m,h))\\&s.t.\ \int_0^1 u^N(W-I-mq^N+s^N\pi,q^N,h)dh \geq \overline{U}\end{aligned}$$

Define $\overline{\pi}$ as the equilibrium level of profit that solves this problem. It can then be equivalently rewritten as:

$$\begin{aligned}&\max_{I,m} \int_0^1 u^N(W-I-mq^N+s^N\overline{\pi},q^N,h)dh\\&s.t.\ 2I+(m-MC)E(q^C(W-I,m,h))+(m-MC)E(q^N(W-I,m,h)) \geq \overline{\pi}\end{aligned}$$

Substituting the reservation profit constraint into the objective function transforms this into an unconstrained maximization problem. With some algebraic manipulation, we can derive an endowment of shares in the firm such that this problem is identical to the second-best competitive equilibrium that solves equation 11.[6] Clearly, monopoly may have distributional consequences, if the shares of the firm are not distributed in the prescribed manner, but it remains (second-best) Pareto-optimal in the usual sense.

However, efficiency losses can occur in the case where there are incentives for uninsurance. Define $(I^{N*},m^{N*})$ as the optimal insurance contract that would be offered to type $N$ in the absence of type $C$ consumers. And define $(I^{C*},m^{C*})$ similarly. When both types remain insured, the firm

---

[6] Specifically, this problem is identical to equation 11, provided that $s^N=\dfrac{(m-MC)E(q^N)+I}{(m-MC)\left[E(q^N)+E(q^C)\right]+2I}$, or that each type gets a share of the firm equal to the share of profits it generates for the firm.

offers the contract $(I^{N*}, m^{N*})$. However, this *may* generate fewer profits than the firm could earn by offering $(I^{C*}, m^{C*})$ to the $C$ types alone. In this case, the firm will choose to price the $N$ types out of the market This is the first efficiency consequence of distortionary pooling equilibria. The healthier (or poorer) consumers with less demand for insurance end up uninsured, where they face a linear monopoly price and all its corresponding deadweight losses. However, the $C$ types continue to enjoy their second-best quantity levels. The empirical consequences of health care monopoly for uninsurance have been considered by Town et al (2006). [7]

### C.3 Effects of Market Power on Under-Insurance

When the premium becomes a blunter instrument for surplus-extraction, the firm will shift towards other methods of extracting additional surplus. One method is by pricing out the low-surplus consumers from the market, as shown above. Another is by "copay-discriminating" and charging higher copays for goods that the $C$ types value more. The result in this case may be under-insurance for the $C$ types, in contrast to uninsurance for the $N$ types.

To take a concrete example, sicker or richer consumers may place more value on the use of high-cost specialty drugs, branded drugs, visits to their own physicians who may be out of the insurer's network, and so on. Therefore, while the insurer may not be able to charge these different types different premiums — or even identify them ex ante — it can charge different prices for these substitutable goods.

Suppose there are two totally distinct types of medical care, both with the same marginal cost, $MC$. For simplicity, we assume that the marginal utility of one of the two goods is zero for

[7] Town et al find that hospital market power leads to increases in rates of uninsurance, although the increase in private uninsurance is almost matched by a relatively large increase in public insurance rates. While we have not considered the impact of public safety nets, their presence could mitigate the welfare impact of market power in the presence of pooling equilibria.

each consumer type, or that each consumer type consumes only one type of good.[8] To make this problem distinct from the earlier analysis, we assume that neither type is "priced out" of the insurance market.

The problem at hand is similar to the earlier pooling equilibrium, except that the firm is now able to charge two different copayment rates on the two distinct goods. It thus solves:

$$\begin{aligned}&\max_{I,m^C,m^N} 2I + (m^C - MC)E(q^C(W-I,m^C,h)) + (m^N - MC)E(q^N(W-I,m^N,h))\\ &s.t. \int_0^1 u^N(W - I - m^N q^N + s^N \pi, q^N, h)dh \geq \overline{U}\end{aligned}$$

Using our familiar methods, we can formally demonstrate equivalence between this problem and the second-best competitive problem in 11. The intuition is straightforward. We have relaxed one constraint on the firm. Therefore, the amount of profit it earns on the $N$ types will not fall. To maximize the surplus extractible from the $N$ types, it will charge the premium $I^{N*}$ and the copayment $m^{N*}$ on the low-value goods.[9]

It remains to identify the optimal level of $m^{C*}$, which solves:

$$\begin{aligned}&\max_{m^C} (m^C - MC)E(q^C(W - I^{N*}, m^C, h))\\ &s.t. \int_0^1 u^C(W - I^{N*} - m^C q^C + (1 - s^N)\pi, q^C, h)dh \geq \overline{U}\end{aligned} \tag{12}$$

Employing our now familiar methods of transformation, we can rewrite this problem as:

$$\max_{m^C} \int_0^1 u(W - m^C(q^C - E(q^C)) - MC * E(q) + \left(I^{C*} - I^{N*}\right), q^C, h)dh, \tag{13}$$

[8] The results would easily generalize to the case in which marginal utilities are always positive, but that consumers differ in their relative valuations.

[9] This can be shown more formally by transforming this problem into the second-best competitive problem, following the reasoning above.

This expression is identical to the second-best planner's problem, *except* for the term $I^{C*} - I^{N*} > 0$. For every choice of copayment, therefore, the chronically ill consumer is richer than in the second-best. As a result, the firm will distort the copay upwards, where the degree of distortion depends on the size of $\left| I^{C*} - I^{N*} \right|$.

## D. Innovation

A major reason for monopolies in health care is the use of patents to encourage innovation. While patents improve dynamic efficiency, two well-known sources of dynamic and static inefficiency remain (Shavell and van Ypersele, 1998). First, incentives to invest in research remain inadequate, because monopoly profits are less than the social surplus created by the innovation. Second, patents encourage innovation at the expense of static inefficiency from monopoly loss. Two-part health insurance can solve both these problems in health care markets – it limits static inefficiency by subsidizing medical care, and at the same time delivers social surplus to a monopolist in the form of the extracted premium. Thus, it can produce better dynamic incentives for innovation, even while it decreases the static costs associated with encouraging innovation. The only danger arises not from patent protection, but from failure in the insurance market: if health insurance is inefficiently cheap or over-provided (due to government subsidies, for example), the result will be excessive amounts of innovation (Garber et al., 2006).

### D.1 The Efficient Allocation

It is well-known that competition does not produce socially optimal outcomes with innovation. Therefore, to calculate the efficient allocation we first solve the Pareto problem. We build upon the framework of the basic incomplete information model presented in Section B.1. To augment that model, consider an economy researching a continuum of innovations, indexed over $i \in [0,1]$. For simplicity, suppose that any given consumer benefits from at most one particular innovation. For each value of $i$, there exists a continuum of consumers indexed over their ex post health status, $h \in [0,1]$. Effectively, this economy consists of consumers populating the unit square, $(h,i) \in [0,1]x[0,1]$, defined over health status and demand for a particular innovation.

If innovation $i$ is discovered, consumers of type $i$ enjoy expected utility $\int_0^1 u^i(W - I^0 - mq^*, q^*, h)dh$. Society has access to an insurance technology for each of the discovered innovations: each consumer of type $i$ pays the unconditional insurance premium $I^0$, and receives the copayment $m$ on the newly discovered technology. This insurance scheme is required to break even.

If innovation $i$ is not discovered, consumers of type $i$ consume zero units of the medical remedy and have no demand for insurance as a result. To compartmentalize the risks from innovation and health shocks, we assume that innovations are discovered before the consumer has to make an insurance purchase decision. There is no way to insure against the failure to discover innovations for those consumers who need them.

The more society invests in research and development, the more innovations are discovered. Define $0 \le \rho(r) \le 1$ as the fraction of innovations discovered; in particular, all innovations for which $i \le \rho(r)$ are discovered. The Pareto problem can thus be written as:

$$\begin{aligned} &\max_{I,m,r} \rho(r)\int_0^1 u(W - I^0 - mq, q, h)dh + (1-\rho(r))\int_0^1 u(W,0,h)dh \\ &s.t.\ \rho(r)\left[I^0 + (m - MC)E(q)\right] \ge r \end{aligned} \tag{14}$$

By displacement, we can write this as:

$$\max_{I,m,r} \rho(r)\int_0^1 u(W - m\left(q - E(q)\right) - \frac{r}{\rho(r)}, q, h)dh + (1-\rho(r))\int_0^1 u(W,0,h)dh \tag{15}$$

It is straightforward to show that innovations, when discovered, are utilized (second-best) efficiently, where $m < MC$. Moreover, research is also efficiently performed, so as to maximize social welfare.

### D.2 The Monopoly Allocation with Two-Part Health Insurance

Now suppose that innovation is performed by an integrated insurer-producer-innovator, who is a monopolist. This firm has access to two-part health insurance pricing. Define $\pi^i$ as the profits earned by consumer type $i$. Profits are distributed as follows: consumers for whom $i \geq \rho(r)$ receive net distributions of zero; those for whom $i \leq \rho(r)$ share equally in the firm's positive profits. The integrated innovator solves the problem:

$$\begin{aligned}&\max_{I,m,r} \rho(r)\big(I + (m - MC)E(q)\big) - r \\ &s.t.\ \int_0^1 u(W - I - mq + \pi^i, q, h)dh \geq \int_0^1 u(W, 0, h)dh,\ \forall i \leq \rho(r)\end{aligned} \tag{16}$$

Defining $\Pi$ as the innovator's total profits in equilibrium and adding a constant term to the objective function, we can rewrite this as:

$$\begin{aligned}&\max_{I,m,r} \rho(r)\int_0^1 u(W - I - mq + \pi^i, q, h)dh + (1 - \rho(r))\int_0^1 u(W, 0, h)dh \\ &s.t.\ \rho(r)\big(I + (m - MC)E(q)\big) - r \geq \Pi\end{aligned} \tag{17}$$

In equilibrium, $\pi^i$ is equal for all $i \leq \rho(r)$, and $\Pi = \rho(r)\pi^i$. Therefore, we can rewrite this as:

$$\max_{I,m,r} \rho(r)\int_0^1 u\Big(W - m\big(q - E(q)\big) - \frac{r}{\rho(r)}, q, h\Big)dh + (1 - \rho(r))\int_0^1 u(W, 0, h)dh \tag{18}$$

This is identical to the planner's problem in equation 15.

### D.3 Impediments to Efficient Innovation

The analysis above considered an unregulated, unsubsidized, and competitive insurance market. In practice, however, employer-based health insurance premia are implicitly subsidized, because they are tax-exempt. This affects the optimal level of the insurance premium generally, along with the incentive to innovate, but it does not affect the optimal copayment, or static efficiency in the goods market.

If consumers face less than the full price of insurance, monopolists will be able to extract consumer surplus *plus* the value of the premium subsidy. However, monopolists will continue to have incentives to set the co-payment so as to maximize extractible consumer surplus. The result is that premium subsidies or taxes affect dynamic efficiency, but not static inefficiency, which the monopolist has incentives to maintain.

As Garber, Jones, and Romer (2006) have argued, this logic suggests that premium subsidies lead to over-innovation. If the innovator can extract total surplus, *in addition to* the value of the premium subsidy, the return on innovation is too high relative to first-best. The result is too much innovation, but efficient provision of the innovations that exist. Notice that we continue to have the result that two-part pricing erases static losses from monopoly, even in the context of innovation.

Additionally, a more complicated model of the innovation process could also lead to inefficiency. In the standard model used above, innovators ought to appropriate the full value of social surplus. Many analysts have pointed out that patent races, public subsidies, and other imperfections can alter this result, so that innovators ought to receive less than social surplus in the first-best allocation. Others, in contrast, have emphasized how little innovators are able to appropriate.[10] This is a difficult question to resolve in our context, because — outside of the simple model presented above — there are a great many possible models of the innovation process, each with different implications. Depending on the first-best rate of appropriation, access to two-part health insurance pricing may result in inefficiently high profits. This affects the optimal tax-and-transfer policy that should accompany a functioning market for health insurance — the social planner can undo incentives to over-innovate by taxing the profits of successful innovators. Regardless of

[10] For contrasting views in the context of pharmaceuticals, see Garber, Jones, and Romer (2006), compared with Philipson and Jena (2006). In a broader context, see Shapiro (2007), compared with Nordhaus (2004). The analytics of patent races appear in Reinganum (1989).

dynamic incentives, two-part pricing through health insurance continues to ensure static efficiency, although it may require correctives to ensure dynamic efficiency as well.

## E. An Empirical Test of Two-Part Pricing

In this section, we test the key implications of our model. First, the model predicts that insurance facilitates two-part pricing. When ideally implemented, two-part pricing ensures that quantity remains at its socially optimal level, even in the presence of market power. Therefore, in markets where consumers are better insured, we expect to see: (1) Reductions in market power will have less, or even no, positive impact on equilibrium quantity; but (2) reductions in market power will have larger negative impacts on the profits or revenues of oligopolistic firms.

We test these predictions in the market for pharmaceuticals, where patent expiration provides a transparent change in market power for the market in a particular molecular entity. The conventional wisdom is that patent expirations *always* increase the quantity of drugs sold. We predict first that the size of this increase is inversely related to insurance, and that molecules that are nearly fully insured should exhibit no increase in quantity at patent expiration. The empirical counterpart of the second prediction is that the size of the decline in *branded* drug revenues will be larger for better-insured molecules.

### E.1 Data

We use the IMS Generic Spectra database for this analysis. These data represent 101 unique molecules, whose patents expired between 1992 and 2002. For each molecule, we use up to 5 years of monthly data, which span the interval from 3 years prior to 2 years after patent expiration. The monthly data include both revenues and quantities. Drug quantity is available in grams. Revenue and quantity data are collected at the retail level (through both retail and hospital pharmacies). IMS then adjusts the revenue data, using proprietary estimates of drug mark-ups, to estimate the implied wholesale revenue. We drop 6 drugs with missing drug sales data.

We use two sources of information to determine the insurance status of drugs. First, we determine whether the drug is primarily consumed and available in hospitals. These hospital drugs are typically covered by medical insurance rather than prescription drug insurance. Since medical insurance is both more common and more generous than prescription drug insurance hospital drugs enjoy more generous insurance coverage than other drugs. We identified 16 hospital drugs in our data. Second, we merged the IMS databases with the 1996 to 2002 Medical Expenditure Panel Survey (MEPS). MEPS contains detailed information on prescription drug use and insurance coverage of a nationally representative population. For each drug we can calculate the share of expenses paid by the patient (out of pocket expenses) and share of expenses paid by the insurer. We use the share of expenses paid by the insurer one year before patent expiration as a measure of insurance coverage for the drug. However, since MEPS data is available only from 1996 onwards, this measure is only available for the 43 drugs whose patent expired in 1997 or later.

### E.2 Empirical Strategy

The identification strategy uses patent expirations as an exogenous source of variation in market power. We test whether the decline in market power after patent expiration leads to a larger increase in equilibrium quantity for drugs with more uninsured patients, and vice-versa. As discussed above we use two different measures of insurance prevalence in the market for a drug: (1) whether the drug is a hospital drug, (2) the percent of drug expenditures borne by insurers. The empirical models we estimate are:

$$LnQ = a + bPatExp + poly(month) + d + \varepsilon \tag{19}$$

$$LnQ = a_1 + b_1 PatExp + c_1 PatExp * Ins + poly(month) + poly(month) * Ins + d + \varepsilon \tag{20}$$

$$LnQ = a_2 + b_2 PatExp + c_2 PatExp * Hosp + poly(month) + poly(month) * Hosp + d + \varepsilon \quad (21)$$

$LnQ$ is the log of total grams of the drug sold in a given month. $PatExp$ is an indicator variable for the months after patent expiration. $Ins$ measures the share of drug expenditures borne by insurers the year before patent expiration. $Hosp$ is an indicator variable for whether the drug is a hospital product. $poly(month)$ is a cubic polynomial in months since patent expiration, and $d$ is a drug fixed-effect.

The first equation models the increase in the quantity of the drug sold after patent expiration. The cubic in month controls for trends in drug sales related to the life-cycle of the drug. The drug fixed-effects absorb time-invariant differences across drugs. The key coefficient of interest is $b$, which measures the percentage change in total quantity due to increased competition following patent expiration. The next equation estimates whether the change in total quantity following patent expiration differs for drugs with greater share of expenditures borne by insurers. A negative estimate for the coefficient $c_1$ would be consistent with our model and would indicate that insured drugs experience a smaller increase in quantity following patent expiration. The last equation models whether the change in total quantity following patent expiration is different for hospital drugs. Again, a negative estimate for coefficient $c_2$ would be consistent with our model and would indicate that hospital drugs (which enjoy more generous insurance) experienced a smaller increase in quantity following patent expiration.

Next we re-estimate the above models with the revenue of branded drugs sold as the outcome variable. These models test whether branded sales decrease after patent expiration and whether the extent of the decrease is related to insurance coverage for the drug.

### E.3 Empirical Results

The results from our empirical tests are presented in Table 1.

**Table 1: Results of Empirical Tests**

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---|---|---|---|---|---|---|
| | log total grams | | | log brand dollars | | |
| Patent Expired | 0.067 | 0.093 | 0.249 | -0.175 | -0.148 | -0.018 |
| | [0.020]*** | [0.021]*** | [0.105]** | [0.029]*** | [0.029]*** | [0.113] |
| Hospital Product*Patent Expired | | -0.158 | | | -0.165 | |
| | | [0.048]*** | | | [0.087]* | |
| Share Insured*Patent Expired | | | -0.344 | | | -0.492 |
| | | | [0.135]** | | | [0.175]*** |
| Constant | 12.323 | 12.322 | 12.572 | 15.968 | 15.968 | 16.272 |
| | [0.020]*** | [0.020]*** | [0.033]*** | [0.024]*** | [0.024]*** | [0.040]*** |
| Observations | 5002 | 5002 | 2488 | 5002 | 5002 | 2488 |
| Number of drugs | 95 | 95 | 43 | 95 | 95 | 43 |

Notes: Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

For the average molecule, Model 1 shows that the total quantity of drugs sold after patent expiration increases by 6.7% after patent expiration. Model 2 shows that hospital drugs, which typically enjoy more generous insurance coverage, experienced a much smaller increase in total quantity of drugs following patent expiration. In fact, the results suggest that hospital drugs experienced no statistically significant change in quantity after patent expiration: the point-estimate is -6.5%. The results from Model 3 suggest that drugs with a larger share of expenses borne by the insurer experienced a smaller increase in quantity of drugs sold after patent expiration.

Model 4 shows that branded drug sales decline by 17.5% after patent expiration. The results from Model 5 and Model 6 show that the decline in branded sales is much larger for hospital drugs and for drugs with a larger share of expenses borne by the insurer. These results are consistent with our model that predicts that reduction in monopoly power should have larger effects on branded revenues for insured drugs.

One threat to the validity of our results is that insured and uninsured drugs might have different entry costs either due to the underlying differences in production technology or regulatory hurdles. For example, if insured products have higher entry costs then these drugs might enjoy less competition from generics after patent expiration; consequently, patent expiration would lead to lesser increase in quantity for insured drugs. To test this alternate explanation, we re-estimate our models with generic market share as the outcome variable. Under this alternate hypothesis, the increase in generic share following patent expiration would be much smaller for hospital drugs and for drugs with a larger share of expenses borne by the insurer. Results from this analysis are presented in Table 2 below.

**Table 2: Testing for differences in generic share after patent expiration**

| | Model 1 | Model 2 | Model 3 |
|---|---|---|---|
| | Generic share in quantity | | |
| Patent Expired | 0.196 | 0.202 | 0.222 |
| | [0.013]*** | [0.015]*** | [0.049]*** |
| Hospital Product*Patent Expired | | -0.032 | |
| | | [0.035] | |
| Share Insured*Patent Expired | | | 0.061 |
| | | | [0.075] |
| Constant | 0.087 | 0.087 | 0.087 |
| | [0.006]*** | [0.006]*** | [0.009]*** |
| Observations | 5002 | 5002 | 2488 |
| Number of drugs | 95 | 95 | 43 |

Notes: Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Model 1 shows that patent expiration leads to a 19.6 percentage point increase in generic market share. Models 2 and 3 show that both hospital drugs and drugs with generous insurance coverage experience a similar increase in generic market share following patent expiration. Thus, there is no prima facie evidence that insured drugs have higher entry costs.

Overall the results from the empirical tests indicate that insurance facilitates two part pricing. These empirical results show that health insurance can significantly reduce the deadweight loss from monopoly pricing of pharmaceuticals.

## F. Conclusions and Implications for Policy

Two-part pricing is well-known as a solution to the deadweight loss from monopoly, but it is frequently impractical. In health care markets, the observed structure of insurance contracts provides a means for achieving the efficient outcomes associated with two-part pricing. While it is not a panacea for informational problems in the insurance market, it can be a useful remedy to static deadweight losses from monopoly, as we have shown. The evidence suggests that the penetration of health insurance lowers the deadweight loss associated with market power, and facilitates the efficient extraction of rents by monopolists. In other words, a well-functioning insurance market transforms the problem of market power from one of efficiency into one of distribution.

A review of health care markets in the late 1990's highlights three interrelated trends: an increase in managed care as method of financing and delivering care; horizontal consolidation within insurer, hospital and physician markets and blurring of the vertical distinctions between these markets (Gaynor and Haas-Wilson, 1999; 2002). Our analysis has important implications for analyzing the potential consequences of each of these trends.

First, our analysis suggests that the recent increase in horizontal consolidation and market power of health care providers may or may not reduce social welfare; the impacts depend on the structure and functioning of the relevant health insurance markets. Moreover, if welfare reductions do occur, they will tend to take the form of uninsurance or under-insurance, rather than simple and direct quantity-restrictions by monopolists.

Second, our analysis suggests that the rise in managed care and vertical integration of health care markets experienced in the 1990's provides unique benefits to society. From a positive point of view, our analysis suggests that vertical integration in health care may be motivated in part by the improved ability of an integrated firm to price-discriminate. This can help to explain why some pharmaceutical companies have chosen to invest in pharmacy benefit managers, and why health-maintenance organizations integrate health-care provision with insurance.

Innovation is of obvious importance in health care markets. Our analysis shows that two-part health insurance pricing has important implications for dynamic incentives. Well-functioning insurance markets may help patent monopolists to extract the maximum amount of consumer surplus associated with their inventions, without restricting quantity. Our empirical analysis of pharmaceutical patent expirations suggests that this may be an important force in innovative health care markets. In well-insured markets, longer patents may have smaller social costs in terms of deadweight loss from monopoly. However, this does not speak to the dynamically optimal level of profits and innovation, which may be affected by patent races, government subsidies to the insurance market, or other factors that can lead to excessive profits accruing to innovators.

The design of public health insurance often considers the trade-offs among optimal risk-bearing, moral hazard, and adverse selection. However, our analysis suggests that it ought to consider how the two-part health insurance contract can best maximize social surplus. An optimally designed public health insurance scheme would set co-payments at or below marginal cost (depending on the extent of moral hazard). The division of resources among consumers can then be determined by the schedule of premia, which allows the government to extract as much or as little consumer surplus as it chooses.

# Appendix

## A. Incomplete Market Power

So far, we have considered the case of pure uncontested monopoly. Many health care markets are better approximated by monopolistic competition. For example, two drug companies might hold patents on different drugs that treat the same disease. Doctors may build unique relationships with their patients, who develop a preference for one physician over another. Patients may prefer to go to a hospital that is closest to their home. All these factors can create product differentiation in the minds of consumers. Market power results, but it is incomplete. In this section, we add monopolistic competition to the moral hazard information structure.

Monopolistic competition changes the distribution of resources relative to complete monopoly, but leaves intact the result that monopolistic competitors choose quantity so as to maximize extractible surplus. A monopolistic competitor must be mindful that her customers can defect to the other firm. This limits the amount of surplus available for extraction. However, conditional on consumer purchases from her, she will continue to set quantity so as to maximize their surplus.

To distill the key ideas, suppose we have two monopolistic competitors—A and B—and two kinds of consumers, with one strictly preferring A, but the other strictly preferring B. Both products have the same marginal cost of production. The firms are integrated in the sense that they both produce their goods and provide insurance contracts over them. Further, as with most spatial models of product differentiation, assume that consumers must choose to use one or the other of the products, but not both—these might be different drugs, physicians, or hospitals, which cannot be easily used with those of rivals. Define $u^A(c,q,h)$ as utility for consumers who prefer $A$ and define $u^B(c,q,h)$ similarly. If a consumer uses the "wrong" good, she derives utility $u^i(c,\delta q,h)$, where $\delta < 1$. Since each consumer can only consume one of the goods, we can assume without loss of generality that insurers provide two insurance contracts—one that provides good $A$ and one that provides good $B$.

### A.1 The Second-Best Efficient Allocation

Clearly, the efficient allocation provides each consumer with her preferred good, and its associated insurance contract. Goods are sold at marginal cost to the insurer. Each contract maximizes the utility of the consumer, subject to the break-even constraint of the insurer. As before, the insurer knows the quantity of good $j$ demanded by a consumer of type $j$ in health state $h$, given the co-payment and income, according to $q^j(W-I,m,h)$.

The optimal contract for the type $j$ consumer maximizes:

$$\begin{aligned}
&\max_{I^j, m^j \le MC} \int_0^1 u^j(W - I^j - m^j q^{j*}, q^{j*}, h)dh \\
&s.t.\ I^j + (m^j - MC)E(q^{j*}) \ge 0 \\
&and\ q^{j*} = q^j(W - I^j, m^j, h)
\end{aligned} \tag{22}$$

This problem is identical to the earlier case of moral hazard, and has a similar solution, characterized by:

$$\begin{aligned}
&[I]: Eu_W^j = \mu\left(1 - (m^j - MC)E(q_W^j)\right) \\
&[m]: \int_0^1 \frac{q^j}{E(q^j)} u_W^j(W - I^j - m^j q^j, q^j, h)dh = \mu\left(1 - (m^j - MC)\frac{-E(q_m^j)}{E(q^j)}\right)
\end{aligned} \tag{23}$$

The insurer sets a co-payment below marginal cost, in an effort to provide some insurance.

### A.2 Equilibrium with Monopolistic Competition

The key difference between monopolistic competition and the earlier case of pure monopoly is in the consumer's reservation utility level. The pure monopolist had only to guarantee the consumer as much utility as she could derive without consuming any medical care goods. The monopolistic competitor, on the other hand, has to guarantee the utility she could derive from the competitor's contract. As with most models of oligopoly, this reservation utility level depends on the absence, presence, and nature of strategic behavior between competitors. However, this does not affect the marginal valuation of goods, only the level of profit earned by the firm. The division of resources among the two firms and the set of consumers have no impact on efficiency. Indeed, if type $j$ consumers own firm $j$, all profits extracted are returned to the consumers from which they were taken. The result is the same equilibrium observed under pure competition.

Without loss of generality, we will demonstrate this reasoning for firm $A$. Define $q^{BA}(W - I^B, m^B, h)$ as the amount of good $B$ that consumer $A$ will use when offered the good $B$ insurance contract. Firm $A$ then solves:

$$\begin{aligned}
&\max_{I^A, m^A} I^A + (m^A - MC)E(q^A) \\
&s.t. \int_0^1 u^A(W - m^A q^{A*} - I^A + \pi^A, q^A, h)dh \geq \\
&\int_0^1 u^A(W - m^B q^{B0} - I^B + \pi^A, q^B, h)dh \\
&q^{A*} = q^A(W - I^A, m^A, h) \\
&q^{B0} = q^{BA}(W - I^B, m^B, h)
\end{aligned} \tag{24}$$

The decisionmaking of the other firm only enters insofar as it affects the consumer's reservation utility level. If consumers own their respective firms, this will not even affect the distribution of resources.

Arguing as we did in the case of moral hazard, define $\overline{\pi}^A$ as the optimal level of profit that solves the firm's problem. The problem in 24 can be equivalently written as:

$$\begin{aligned}
&\max_{I^A, m^A} \int_0^1 u^A(W - m^A q^{A*} - I^A + \overline{\pi}^A, q^A, h)dh \\
&s.t.\ I^A + (m^A - MC)E(q^A) \geq \overline{\pi}^A \\
&q^{A*} = q^A(W - I^A, m^A, h)
\end{aligned} \tag{25}$$

The displaced version of this problem is identical to the displaced version of the competitive problem in 22. This demonstrates that monopolistic competition produces the same allocation as pure competition.

### A.3 Two-Sided Market Power

Outcomes may not be perfectly efficient if both the insurance and provider sides are characterized by incomplete market power. In the case of bilateral monopoly, both sides continue to have incentives to maximize consumer surplus, and then divide it between themselves. However, imperfect competition on both sides can create unique distortions. For example, there may be strategic incentives for exclusive dealing, which is often observed in pharmaceutical markets. Specific insurers award low co-payments to a few drugs in a specific therapeutic class. This then allows them to extract more favorable terms from the manufacturer, because they can promise higher volume sales (cf, Ellison and Snyder, 2003).

We simplified the problem by considering insurance policies for a single innovation. This simplification sacrifices no generality when the insurance market is perfectly competitive. Even if each insurance policy covered a large number of possible therapies, a perfectly competitive insurance industry with a large number of insurers could offer a variety of policies that covered the preferences of every consumer. However, with a limited number of insurers, but a large number of therapies, it is possible that some consumers might prefer a set of therapies that is not well-covered.

This point can be demonstrated with a simple example. Suppose there are ten therapies to treat a single disease, and two innovators — innovator A sells 9 of these therapies, while innovator B sells only one. There is a single insurer, and ten consumers. Each consumer derives $100 of surplus from the therapy she prefers: Nine consumers prefer one of A's therapies, while the tenth prefers B's therapy. Suppose innovator A demands an exclusive contract with the insurer. This is a credible demand if all $90 of consumer surplus is extracted, and if the innovator gives the insurer $15 of this surplus. Innovator B cannot match the offer. The result is that utilization of B's therapy is inefficient, because the patient preferring B can only buy it directly, and not through an insurance policy. This leads to the typical monopoly problem, and the under-provision it commonly implies.

This suggests that market power in the insurance industry may be the root cause of inefficient utilization, rather than market power on the provider/innovator side. This also suggests that inefficiency in the insurance industry will cut against the ability of insurance to produce perfectly efficient outcomes. Nonetheless, in actual practice, private insurance contracts provide significant price reductions on a large number of therapies and treatments, even though the largest price reductions might be reserved for a few "preferred" drugs or providers. Even so, the actual price discounts observed lead to significant reductions in deadweight loss, and improvements in efficiency.

## B. Adverse Selection

To model adverse selection, suppose that consumers are heterogeneous ex ante. There are chronically ill patients (type $C$), and not chronically ill patients (type $N$). Firms cannot observe consumer types. Define $\mu^C(h)$ and $\mu^N(h)$ as the distributions of chronically ill and not chronically ill people. The health distribution for the chronically ill is assumed to dominate the other in the first-order stochastic sense. An insurance contract is an ex ante insurance premium $(I)$, coupled with an ex post copayment $(m)$.

### B.1 The Competitive Solution

A pooling equilibrium is not possible for the usual reasons (Rothschild and Stiglitz, 1976): given any putative pooling equilibrium, there is always a profitable contract that attracts only the low-risk insureds. Therefore, if an equilibrium exists, it must be a separating equilibrium. As such, the competitive insurance industry chooses two contracts that maximize the welfare of each type of agent, subject to incentive compatibility constraints (ensuring the contracts are chosen by the correct agents), and break-even constraints. The contract $(m^C, I^C)$ for the chronically ill solves:

$$
\begin{aligned}
&\max_{m^C, I^C} \int_0^1 u(W - I^C - m^C q^C, q^C, h)\mu^C(h)dh \\
&s.t. \\
&[\gamma]: \int_0^1 u(W - I^N - m^N q^N, q^N, h)\mu^N(h)dh \\
&\geq \int_0^1 u(W - I^C - m^C q^C, q^C, h)\mu^N(h)dh \\
&[\beta]: I^C + \int_0^1 q^C \mu^C(h)dh(m^C - p) \geq 0 \\
&q^N \equiv q(W - I^N, m^N, h), q^C \equiv q(W - I^C, m^C, h)
\end{aligned} \tag{26}
$$

This problem has the following first-order conditions:

$$
\begin{aligned}
&[I]: E_C(u_W) - \gamma E_N(u_W) = \beta(1 - E_C(q_W)(m^N - p)) \\
&[m]: E_C(u_W \frac{q^C}{E_C(q^C)}) - \gamma E_N(u_W \frac{q^C}{E_C(q^C)}) = \beta(1 - \frac{E_C(-q_m)}{E_C(q_C)}(m^C - p))
\end{aligned} \tag{27}
$$

Notice that if the incentive constraint fails to bind, these first-order conditions are identical to the second-best equilibrium with moral hazard.

This observation reveals how the adverse selection equilibrium is affected by the introduction of moral hazard. In the absence of moral hazard, full insurance is the benchmark outcome. Full insurance is never incentive-compatible, because high-risk consumers always prefer the full insurance contract offered to the lower-risk, lower-cost consumers. This explains why, in the standard Rothschild-Stiglitz setting, adverse selection always impacts outcomes. In this case, however, the second-best moral hazard contracts may sometimes be incentive-compatible. Suppose, for example, that the second-best contract involves a very high copayment for the low-risks, because they have a highly elastic demand and relatively little insurable risk. If so, it is possible that the high-risk insureds would prefer their own second-best contract to that offered to the low-risks. In this event, adverse selection would have no impact, because incentive compatibility emerges of its own accord, due to moral hazard. This would leave us with the moral hazard equilibrium outlined above.

If, however, the second-best contracts are not incentive-compatible, we obtain the typical Rothschild-Stiglitz solution in which the high-risk consumers receive their second-best contract, but the low-risk consumers receive something worse than their second-best.

The indirect utility conferred by a specific contract is defined by $v^C(I,m)$ and $v^N(I,m)$ for the chronically ill and not chronically ill patients, respectively; these are defined as follows.

$$v(I,m) \equiv \max_q \int_0^1 u(W-I-mq(W-I,m,h), q(W-I,m,h),h)\mu(h)dh \qquad (28)$$

We impose two assumptions that make this environment similar to the Rothschild-Stiglitz one. First, the chronically ill are willing to pay more for a given change in the copayment rate, in the sense that:

$$-\frac{dI}{dm}\Big|_{v^C} > -\frac{dI}{dm}\Big|_{v^N} \qquad (29)$$

This is the typical "single-crossing" property from Rothschild and Stiglitz's (1976) analysis of adverse selection.[11] Second, a given change in the co-payment rate has a bigger impact on a firm's profits, so that:

$$\begin{aligned} -\frac{dI}{dm}\Big|_{\pi=0} &= \frac{E(q)+(MC-m)E(q_m)}{E(q_W)(MC-m)} \\ -\frac{dI}{dm}\Big|_{\pi^C=0} &> -\frac{dI}{dm}\Big|_{\pi^N=0} \end{aligned} \qquad (30)$$

Figure 1 illustrates the separating equilibrium in $(I,m)$-space. The curves $Z^N$ and $Z^C$ represent the zero-profit curves for the not chronically ill and chronically ill, respectively. $v^C$ is the indifference curve for the chronically ill tangent to the zero-profit line — this represents the optimal (i.e., second-best) contract that is possible under moral hazard. Observe that if the second-best

---

[11] $-\frac{dI}{dm} = \frac{E(u_w q)}{E(u_w)}$. First-order stochastic dominance implies that the numerator is higher for the chronically ill. We assume this effect outweighs the fact that the marginal utility of wealth may also be higher for the chronically ill.

contract for the not chronically ill falls on the curve segment $A$, there is no adverse selection problem, because both second-best contracts are incentive-compatible.

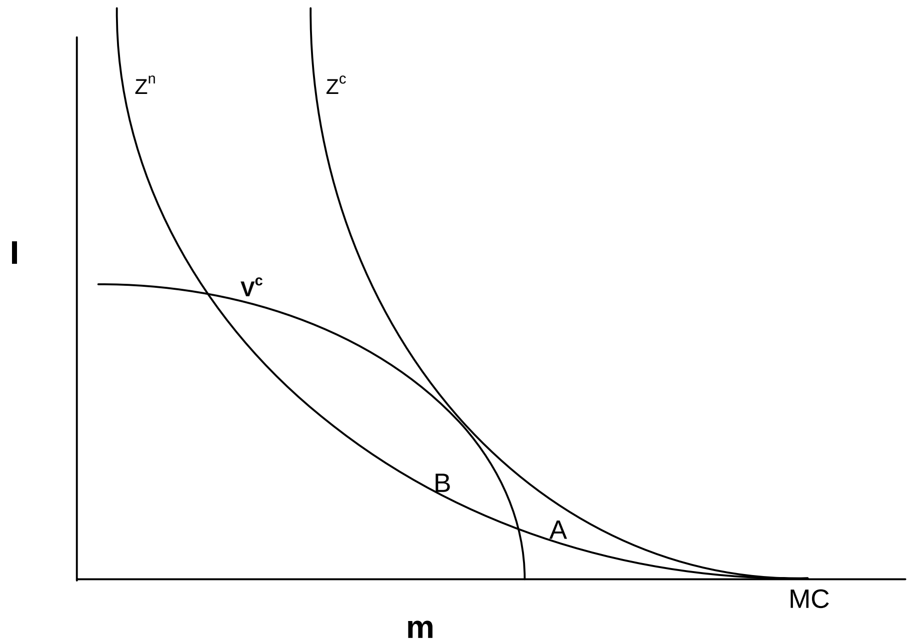


**Figure 1: Equilibrium with adverse selection and moral hazard.**

Now consider the case where adverse selection has an impact: if the second-best contract for type $N$ falls on the curve segment $B$. In this case, the chronically ill will receive their second-best contract, while the other type will receive the contract at the intersection of $v^C$ and $Z^N$.

### B.2 Equilibrium with Two-Part Monopoly Pricing

A monopolist who charges an upfront premium and an ex post copayment maximizes profits subject to reservation utility conditions (i.e., participation constraints) and incentive constraints.

$$
\begin{aligned}
&\max_{m^C, I^C, m^N, I^N} \; I^C + (m^C - MC)\int_0^1 q^C \mu^C(h)dh + I^N + (m^N - MC)\int_0^1 q^N \mu^N(h)dh \\
&s.t. \\
&\int_0^1 u(W - I^N - m^N q^N, q^N, h)\mu^N(h)dh \\
&\geq \int_0^1 u(W - I^C - m^C q^C, q^C, h)\mu^N(h)dh \\
&\int_0^1 u(W - I^C - m^C q^C, q^C, h)\mu^C(h)dh \\
&\geq \int_0^1 u(W - I^N - m^N q^N, q^N, h)\mu^C(h)dh \\
&\int_0^1 u(W - I^C - m^C q^C, q^C, h)\mu^C(h)dh \geq \bar{u}^C \\
&\int_0^1 u(W - I^N - m^N q^N, q^N, h)\mu^N(h)dh \geq \bar{u}^N \\
&q^N \equiv q(W - I^N, m^N, h), q^C \equiv q(W - I^C, m^C, h)
\end{aligned}
\tag{31}
$$

Since this problem is additively separable in $(I^C, m^C)$ and $(I^N, m^N)$, the joint profit-maximization problem is identical to two separate problems, in which the monopolist maximizes profits over each contract. Specifically, the maximization problem in 31 is equivalent to the pair of maximization problems below:

$$
\begin{aligned}
&\max_{m^C, I^C} \; I^C + (m^C - MC)\int_0^1 q^C \mu^C(h)dh \\
&s.t. \\
&\int_0^1 u(W - I^{N*} - m^{N*} q^{N*}, q^{N*}, h)\mu^N(h)dh \\
&\geq \int_0^1 u(W - I^C - m^C q^C, q^C, h)\mu^N(h)dh \\
&\int_0^1 u(W - I^C - m^C q^C, q^C, h)\mu^C(h)dh \geq \bar{u}^C \\
&q^N \equiv q(W - I^{N*}, m^{N*}, h), q^C \equiv q(W - I^C, m^C, h)
\end{aligned}
\tag{32}
$$

$$\begin{aligned}
&\max_{m^N, I^N} I^N + (m^N - MC)\int_0^1 q^N \mu^N(h)dh \\
&s.t. \\
&\int_0^1 u(W - I^{C*} - m^{C*}q^C, q^C, h)\mu^C(h)dh \\
&\geq \int_0^1 u(W - I^N - m^N q^N, q^N, h)\mu^C(h)dh \\
&\int_0^1 u(W - I^N - m^N q^N, q^N, h)\mu^N(h)dh \geq \bar{u}^N \\
&q^N \equiv q(W - I^N, m^N, h), q^C \equiv q(W - I^{C*}, m^{C*}, h)
\end{aligned} \tag{33}$$

As in the moral hazard case, it is straightforward to show that these problems yield Pareto-equivalent allocations to the competitive problems.

Without loss of generality, we show this for the type $N$ contract. To net out distributional effects, we assume that the representative type $N$ consumer holds a claim on all profits that flow from contracts with type $N$ consumers. There may not be a well-defined equilibrium in the case of adverse selection, but for our purposes, it suffices to consider the case where an equilibrium exists. If no equilibrium exists, deadweight loss from monopoly is undefined. Define $\bar{\pi}^N$ as the equilibrium profit associated with the solution to 33. If so, then 33 is identical to a problem in which the firm maximizes consumer utility subject to a reservation profit constraint, and the incentive constraint. This problem will also yield profits equal to $\bar{\pi}$, incentive-compatibility, and utility at least equal to $\bar{u}^N$:

$$\begin{aligned}
&\max_{m^N, I^N} \int_0^1 u(W - I^N - m^N q^N + \bar{\pi}^N, q^N, h)\mu^N(h)dh \\
&s.t. \\
&\int_0^1 u(W - I^{C*} - m^{C*}q^C, q^C, h)\mu^C(h)dh \\
&\geq \int_0^1 u(W - I^N - m^N q^N, q^N, h)\mu^C(h)dh \\
&I^N + (m^N - MC)\int_0^1 q^N \mu^N(h)dh \geq \bar{\pi}^N \\
&q^N \equiv q(W - I^N, m^N, h), q^C \equiv q(W - I^{C*}, m^{C*}, h)
\end{aligned} \tag{34}$$

Substituting the reservation profit constraint into the consumer's objective function yields:

$$
\begin{aligned}
&\max_{m^N, I^N} \int_0^1 u(W - m^N(q^N - E(q^N)) - MC * E(q^N), q^N, h)\mu^N(h)dh \\
&s.t. \\
&\int_0^1 u(W - I^{C*} - m^{C*}q^C, q^C, h)\mu^C(h)dh \\
&\geq \int_0^1 u(W - I^N - m^N q^N, q^N, h)\mu^C(h)dh \\
&q^N \equiv q(W - I^N, m^N, h), q^C \equiv q(W - I^{C*}, m^{C*}, h)
\end{aligned} \tag{35}
$$

This problem is identical to the displaced version of the competitive problem in 26.[12] Therefore, the monopoly allocation is identical to the competitive one.

[12] Under competition, $p = MC$, and $I^C = -(m^C - p)\int_0^1 q^C \mu^C(h)dh$.

## References


Adams, S. (2007). "Health Insurance Market Reform and Employee Compensation: The Case of Pure Community Rating in New York." Journal of Public Economics **91**(5-6): 1119-33.

Arrow, K. J. (1963). "Uncertainty and the Welfare Economics of Medical Care." American Economic Review **53** 941-973.

Bhattacharya, J. and M. K. Bundorf (2005). The Incidence of the Healthcare Costs of Obesity, National Bureau of Economic Research, Inc, NBER Working Papers: 11303.

Bhattacharya, J. and W. B. Vogt (2006). "Employment and Adverse Selection in Health Insurance." National Bureau of Economic Research Working paper 12430. Cambridge, MA.

Crocker, K. J. and J. R. Moran (2003). "Contracting with Limited Commitment: Evidence from Employment-Based Health Insurance Contracts." Rand Journal of Economics **34**(4): 694-718.

Dor, A., S. M. Koroukian and M. Grossman (2004). "Managed Care Discounting: Evidence from the MarketScan Database." Inquiry **41**(2): 159-69.

Dranove, D., M. Shanley and W. D. White (1993). "Price and Concentration in Hospital Markets: The Switch from Patient-Driven to Payer-Driven Competition." Journal of Law and Economics **36**(1): 179-204.

Ellison, S. F. and C. Snyder (2003). "Countervailing Power in Wholesale Pharmaceuticals." MIT Working Paper. Cambridge, MA.

Fang, H. and A. Gavazza (2007). Dynamic Inefficiencies in Employment-Based Health Insurance System Theory and Evidence. National Bureau of Economic Research, Inc, NBER Working Papers.

Garber, A., C. Jones and P. Romer (2006). "Insurance and Incentives for Medical Innovation." Forum for Health Economics and Policy **9**(2): Article 4.

Gaynor, M. and D. Haas-Wilson (1999). "Change, consolidation, and competition in health care markets." Journal of Economic Perspectives **13**(1): 141-64.

Gaynor, M. and D. Haas-Wilson (2002). "Change, Consolidation, and Competition in Health Care Markets." The Heinz School, Carnegie-Mellon Working Paper 2002-19. Pittsburgh, PA.

Gaynor, M., D. Haas-Wilson and W. B. Vogt (2000). "Are Invisible Hands Good Hands? Moral Hazard, Competition, and the Second-Best in Health Care Markets." Journal of Political Economy **108**(5): 992-1005.

Keeler, E. B., G. Melnick and J. Zwanziger (1999). "The Changing Effects of Competition on Non-profit and For-Profit Hospital Pricing Behavior." Journal of Health Economics **18**(1): 69-86.

Levy, R. (1999). "The pharmaceutical industry. A discussion of competitive and antitrust issues in an environment of change." Bureau of Economics/Federal Trade Commission staff report.

Melnick, G. (2004). Hospital Pricing and the Uninsured. Subcommittee on Health of the House Committee on Ways and Means. H. c. o. W. a. Means**:** 6.

Nordhaus, W. D. (2004). "Schumpeterian Profits in the American Economy: Theory and Measurement." National Bureau of Economic Research Working Paper 10433. Cambridge, MA.

Oi, W. Y. (1971). "A Disneyland Dilemma: Two-part Tariffs for a Mickey Mouse Monopoly." Quarterly Journal of Economics **85**: 77-96.

Pauly, M. (1968). "The Economics of Moral Hazard: Comment." American Economic Review **58**(3): 531-537.

Philipson, T. J. and A. B. Jena (2006). "Surplus Appropriation from R&D and Health Care Technology Assessment Procedures." National Bureau of Economic Research Working Paper 12016. Cambridge, MA.

Reinganum, J. (1989). The Timing of Innovation: Research, Development, and Diffusion. Handbook of Industrial Organization. R. Schmalensee, and R. D. Willig, Eds. New York, Elsevier Science. **1:** 850-908.

Rothschild, M. and J. E. Stiglitz (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." Quarterly Journal of Economics **90**(4): 630-49.

Shapiro, C. (2007). Patent Reform: Aligning Reward and Contribution. National Bureau of Economic Research, Inc, NBER Working Papers.

Shavell, S. and T. van Ypersele. (1998). "Rewards versus Intellectual Property Rights." Retrieved December 1998, from SSRN: <http://ssrn.com/abstract=145292> or DOI: 10.2139/ssrn.145292 <http://dx.doi.org/10.2139/ssrn.145292>

Stiglitz, J. E. (1977). "Monopoly, Non-Linear Pricing and Imperfect Information: The Insurance Market." The Review of Economic Studies **44**(3): 407-430.

Town, R., D. Wholey, R. Feldman, et al. (2006). "The Welfare Consequences of Hospital Mergers." University of Minnesota Working Paper. St. Paul.

Zeckhauser, R. (1970). "Medical Insurance: A Case Study of the Tradeoff Between Risk Spreading and Appropriate Incentives." Journal of Economic Theory **2**(1): 10-26.

Zwanziger, J. and G. A. Melnick (1988). "The Effects of Hospital Competition and the Medicare PPS Program on Hospital Cost Behavior in California." Journal of Health Economics **7**(4): 301-20.