Incentivizing efficient utilization without reducing access: The case against cost-sharing in insurance

by Markus Fels

Incentivizing Efficient Utilization Without Reducing Access: The Case Against Cost-Sharing in Insurance

Markus Fels*

13$^{\text{th}}$ July, 2017

Abstract

Cost-sharing is regarded as an important tool to reduce moral hazard in health insurance. Contrary to standard prediction, however, such requirements are found to decrease utilization both of efficient and of inefficient care. I employ a simple model that incorporates two possible explanations - consumer mistakes and limited access - to assess the welfare implications of different insurance designs. I find cost-sharing never to be an optimal solution as it produces two novel inefficiencies by limiting access. An alternative design, relying on bonuses, has no such side effects and achieves the same incentivization.

JEL Classification: D82, I13, I14

Keywords: Moral Hazard, Limited Access, Cost-Sharing, Insurance Rebates

1 Introduction

Moral hazard, specifically ex-post moral hazard, is argued to be one of the main impediments to a well-functioning insurance market. Once insured, individuals no longer pay the full price of the health care that they consume. This increases health care expenditures. Economists have pointed out that this is caused by insurees who overconsume health care as they no longer face its entire cost. Accordingly, the consumption of additional health care by the insured is characterized as welfare-decreasing as it must be valued below cost, for otherwise it would also be consumed in the absence of insurance. In response, economists have proposed partial insurance as a means to deter the consumption of inefficient care.[2] The view that regards moral hazard as entirely inefficient has been challenged, however. Nyman (1999a) points out that a major part of the additional care is consumed only by the insured because it is only affordable with insurance, not because it has little value. If insurance provides access to otherwise unaffordable care, the additional health expenditures of the insured are neither inefficient nor a threat to the well-functioning of an insurance market. On the contrary, the additional expenditures by the insured must then be viewed as the very reason for the existence of this market.[3] Reducing these additional expenditures through partial insurance is then neither necessary nor desirable.[4]

Empirical evidence is robust in showing an effectiveness of partial insurance in reducing utilization (Zweifel and Manning 2000). Contrary to the prediction of inefficient moral hazard, however, the reduction in health spending is not restricted to care that is considered little effective. Instead, insurees react to increased cost by reducing both valuable and less valuable care (Zweifel and Manning 2000; Brot-Goldberg et al. 2015). While this finding remains robust, its interpretation is controversial. Some scholars criticize a welfare analysis based on consumer demand as consumers may lack information to distinguish high-value from low-value care (Rice 1992).[5] Pauly and Blavin (2008) suggest to differentiate cost-sharing arrangements according to

---

[2]For nice overviews of the literature on moral hazard in health insurance, see Zweifel and Manning (2000) and McGuire (2011).

[3]Nyman (1999b) calls this benefit of insurance its access value.

[4]In fact, Fels (2016) shows that deductibles can destroy more than the access value of insurance if affordability constraints matter. In addition, they excessively reduce the actuarial value of insurance by not only reducing the payment given a claim, but also reducing the probability of filing a claim. The latter is due to the fact that, in many insurance markets, benefit payment is conditional on deductible payment. That implies that an insuree is unable to make use of his insurance if he cannot afford to pay the deductible.

[5]Others defend the traditional welfare analysis, insisting that treatment effectiveness is not to be confused with treatment efficiency (Peele 1993). As a highly effective care can also produce high non-monetary cost, the finding

consumers' mistakes in appropriately assessing the quality of care. This idea is essentially reflected in the concept of value-based insurance design as proposed by Fendrick et al. (2001) (see also Fendrick and Chernew (2006)). Still, researchers remain worried that cost-sharing arrangements are an instrument that discriminates only insufficiently between appropriate and inappropriate care (Zweifel and Manning 2000).

In addition, consumer-directed health care, as insurance with strong elements of cost sharing is called, has drawn criticism for reducing access to health care, irrespective of the quality of that care. There is a growing concern that cost-sharing arrangements prevent the consumption of necessary care, in particular among low-income households.[6] This criticism is underscored by evidence that suggests a negative association between insurance coverage and health outcomes for those with low income and poor health (Zweifel and Manning (2000) pp. 442-444, Tamblyn et al. (2001)). An efficient allocation of resources in health care thus requires insurance plans both to deter the consumption of ineffective care (inefficient moral hazard) and to ensure access to effective care (efficient moral hazard). While typical instruments of cost sharing such as deductibles and co-payments seem effective in achieving the first, they seem to fail at the second requirement. The reduction of access to efficient care through cost sharing has - to my knowledge - so far either been entirely ignored or considered as a necessary evil in the insurance literature. This is unfortunate as access concerns are of high societal and political relevance[7] and, as I will point out in this paper, can be appropriately addressed through a different, astonishingly simple insurance design that involves bonuses instead of cost-sharing.[8] The evil turns out to be unnecessary.

In this paper, I analyze different insurance designs to address moral hazard. First, I provide a simple model of insurance that includes both efficient and inefficient moral hazard. The first is a simple consequence of affordability constraints and constitutes the access value of insurance as proposed by Nyman (1999b). The second reflects the well-known idea that full insurance leads

---

of a reduction of effective care does not necessarily contradict the notion that the additional health expenditure of insured individuals is welfare-decreasing.

[6]See e.g. Beck (1974) for early evidence of a stronger response to cost-sharing among lower-income households. In contrast, Chandra et al. (2014) find a response of low-income households that is similar to higher-income groups when measured by demand elasticities. Notably, however, they find that roughly 70% of the spending reduction of low-income households can be attributed to reductions on the extensive margin, i.e. a reduction of utilization to zero. For a recent review of the literature on the relationship between out-of-pocket cost and utilization, see Schokkaert et al. (2017).

[7]For recent media coverage, see Jan (2015) and Pear (2015).

[8]The possibility to use rebates to mitigate moral hazard is already recognized by Rubinstein and Yaari (1983). Here, I point out that such contracts have the additional advantage of not creating access problems.

consumers to seek care even if it is of low value. I show that in this framework, an insurance with bonuses welfare-dominates both full insurance and partial insurance involving cost-sharing. In the following section, I extend the model to allow for consumer mistakes in distinguishing low-value from high-value care. I analyze which insurance design maximizes welfare in a framework that incorporates both of the reasons that have been suggested to explain consumers' indiscriminate response to cost-sharing: access problems and consumer mistakes. Finally, I point out that cost-sharing leads to an additional inefficiency beyond reducing access. It introduces an adverse-selection problem into the insurance market that puts selective pressure on exactly those types that value insurance the most. Again, I show that an insurance, that relies on bonuses instead of cost-sharing, does not produce such a problem of adverse selection. In the final section, I conclude.

2   An Insurance Model with Two Types of Moral Hazard

Suppose a risk-neutral individual faces a probability $\pi \in (0,1)$ of sickness. In case of sickness, health care is available at cost p. The value of care is private information of the consumer at the time of treatment choice. It can confer a high value $V_h > p$, such that treatment is efficient. It can be of low value $V_l < p$, such that it is efficient to abstain from treatment. Ex ante, sickness is associated with a high-value treatment with probability π_h and with a low-value treatment with probability π_l, such that $\pi_h + \pi_l = \pi$. The individual's budget x at the time of need is a random variable from an ex-ante perspective. Let $F(x)$ denote the cumulative distribution function over x, and by $\hat{x} = \int x dF(x)$ the expected wealth at the time of need. Assume a simple additively-separable utility function $u(c,k) = c + k$ where c denotes utility from consumption and k denotes utility from receiving care. W.l.o.g. I normalize k to zero both in the state when no care is needed and in the state when care is needed but not received.

Then the utility from remaining uninsured is given by

$$u_0 = (1 - \pi_h)\hat{x} + \pi_h \left[\rho \mathbb{E}\left[x | x < p \right] + (1 - \rho)(V_h + \mathbb{E}\left[x | x > p \right] - p) \right] \tag{1}$$

$$= \hat{x} + \pi_h(1 - \rho)(V_h - p), \tag{2}$$

where $\rho = F(p)$ denotes the probability of not being able to afford the cost of care p. When remaining uninsured and thus having to pay the full cost of care out of pocket, the consumer decides to receive treatment only if it is of high-value. Yet, even if treatment has high value, the

4

consumer may not be able to afford treatment when his budget falls below p.

Consider an insurance that fully covers treatment cost and is available at a premium w. As the price, that the consumer needs to pay to receive treatment, is reduced to zero, care is consumed irrespective of value. Fully insuring then yields a utility

$$u_{full} = \hat{x} + \pi_h V_h + \pi_l V_l - w. \tag{3}$$

Full insurance leads to the consumption of care whenever sick, even if care is of low value. However, consumption of high-value care is no longer confined to the case when care is affordable. Calculating the premium, that leaves the consumer indifferent between full and no insurance, yields the maximum willingness-to-pay for for full insurance:[9]

$$\bar{w}_{full} = \pi p + \pi_h \rho(V_h - p) - \pi_l(p - V_l). \tag{4}$$

Subtracting the expected cost of full insurance $c_{full} = \pi p$, allows to derive the net surplus of full insurance:

$$s_{full} = \bar{w}_{full} - c_{full} = \pi_h \rho(V_h - p) - \pi_l(p - V_l). \tag{5}$$

The net surplus of insurance consists of two parts. The first part reflects the access value of insurance as described by Nyman (1999b). Insurance is valuable (even to a risk-neutral individual) as it helps to overcome affordability constraints that prevent consumers from receiving efficient care. The second part reflects the familiar problem of full insurance leading to the consumption of inefficient care. It is important to point out that both parts embody a form of moral hazard, as both relate to the consumption of additional care by the insured. The first part constitutes efficient moral hazard as insurance allows consumers to consume efficient care that is otherwise unaffordable, and the second part constitutes inefficient moral hazard, as insurance makes consumers seek care even if it is of low value. Importantly, inefficient moral hazard, if large enough, can reduce the surplus below zero, leaving no gains from trade in the market. Accordingly, economists have proposed to impose cost-sharing as a means to deter the consumption of ineffective care, thereby

---

[9]Note that I abstract from affordability constraints with respect to the premium w. This is a deliberate modeling choice. The model is used to determine the maximum willingness-to-pay for insurance in order to determine the net welfare implications of different insurance designs. Affordability constraints with respect to the premium mean that actual demand no longer appropriately reflects the value that the consumer derives from insurance.

eliminating inefficient moral hazard.

Suppose that insurance no longer fully covers medical expenses but specifies a deductible $d < p$.[10] That is, insurance no longer reduces the price of care for the insuree to zero, but to d. If d is set sufficiently high such that $V_l = d < V_h$, the insuree refrains from seeking care if it is of low value, and seek care if it is of high value. Buying such partial coverage then results in a utility of

$$u_p = (1 - \pi_h)\hat{x} + \pi \left[\delta(\mathbb{E}\left[x|x < d\right]) + (1 - \delta)(V_h + \mathbb{E}\left[x|x > d\right] - d)\right] - w \tag{6}$$

$$= \hat{x} + \pi(1 - \delta)(V_h - d) - w, \tag{7}$$

where $\delta = F(d)$. The willingness-to-pay for and the cost of insurance of partial insurance are then given by

$$\bar{w}_p = \pi_h \left[(1 - \delta)(p - d) + (\rho - \delta)(V_h - p)\right], \tag{8}$$

$$c_p = \pi_h(1 - \delta)(p - d). \tag{9}$$

This results in a surplus of

$$s_p = w_p - c_p = \pi_h(\rho - \delta)(V_h - p). \tag{10}$$

The nonnegativity of the surplus of partial insurance shows that deductibles are able to deter the consumption of ineffective care as predicted. This comes at the cost of confining the consumption of effective care to the state in which the deductible is affordable. The model thus encompasses limited access as a possible explanation for the indiscriminate reduction of utilization that is observed. Essentially, deductibles are too effective in reducing care utilization thereby trading one inefficiency - inefficient moral hazard - for another - the reduction of access to valuable care. Importantly, if $\delta \approx \rho$, the entire surplus of insurance vanishes. If affordability constraints are sufficiently severe such that it becomes equally unlikely for the consumer to afford the deductible as it is to afford the complete cost of care, the entire reason for insurance purchase vanishes. If access motives underlie insurance purchase, cost-sharing amounts to throwing out the baby with the bathwater.

---

[10]Note that if medical expenses are lumpy, there is no need to distinguish between different forms of cost-sharing. d simply specifies the total amount of money the insuree has to spend to seek care, irrespective of whether this is the result of deductibles, co-payments, other fees, or a combination of them. For convenience, I will simply refer to the total amount as deductible.

Suppose that, instead of requiring a deductible payment in case of care consumption, insurance fully covers treatment cost, but promises a bonus payment (or rebate) r in case that care is not consumed. If this no-claim bonus is of appropriate size, $V_l = r < V_h$, the consumer prefers to receive the bonus in case that care is of low value and prefers to seek treatment in case that treatment is of high value. This results in the utility

$$u_r = \hat{x} + (1 - \pi_h)r + \pi_h V_h - w. \tag{11}$$

Calculating willingness-to-pay, expected cost, and net surplus yields

$$\bar{w}_r = (1 - \pi_h)r + \pi_h(p + \rho(V_h - p)), \tag{12}$$

$$c_r = (1 - p_h)r + \pi_h p, \tag{13}$$

$$s_r = \pi_h \rho(V_h - p). \tag{14}$$

Similar to insurance with cost-sharing, rebate insurance limits consumption of care to high-value care. However, in strong contrast to cost-sharing, rebates do not limit the access to high-value care. They are thus able to successfully deter consumption of low-value care, while, at the same time, not inhibiting reception of high-value care. In this way, rebate insurance is able to achieve the first-best outcome. This comes at higher insurance cost of $c_r = (1 - p_h)r + \pi_h p$, as insurance benefits are no longer confined to only paying for treatments. The net welfare surplus of rebate insurance, however, is strictly larger than the surplus of insurance with cost-sharing whenever there is a positive chance that the deductible is unaffordable, $\delta > 0$.

Proposition 1. *If consumers are perfectly informed, rebate insurance is able to fully eliminate inefficient moral hazard while protecting access, thereby achieving the first best.*

The result contradicts the notion that access reduction is a necessary evil if we want to incentivize efficient utilization of health care by consumers. There is indeed a way in which we can achieve the same incentivization without reducing the consumption of efficient care by putting up access barriers. In addition, the larger surplus of rebate insurance can straightforwardly explain evidence that suggests a preference for rebates over deductibles as observed by Johnson, Hershey, Meszaros, and Kunreuther (1993) (see also (Kunreuther, Pauly, and McMorrow 2013), pp. 118-119).[11] It is, however, important to recognize that rebate insurance requires premium

---

[11]Johnson, Hershey, Meszaros, and Kunreuther (1993) propose an explanation based on different frames being

payments that are substantially larger than insurance relying on cost-sharing. Hence, affordability constraints are likely relevant at the time of insurance purchase. I abstract from these in the formal analysis as these can be appropriately addressed through subsidizing insurance purchase without distorting utilization incentives.[12]

3 Consumer Mistakes

Consumer mistakes have been proposed as an explanation for the observed indiscriminate response of consumers to cost-sharing arrangements (Rice 1992; Pauly and Blavin 2008). Consumers may reduce both the consumption of effective and ineffective care in response to cost-sharing if they have problems distinguishing the two. In this section, I augment the model of the previous section to allow for consumer mistakes. The augmentation then incorporates both explanations - access reduction and consumer mistakes - in a simple model that allows to compare the welfare implications of different insurance regimes.

Suppose that consumers make mistakes in their assessment of whether a treatment is of high or low value. Assume that a consumer wrongly assigns a high value to a low-value treatment with probability $\alpha \in [0, 1]$, and wrongly assigns a low value to a high-value treatment with probability $\beta \in [0, 1]$. I call the first mistake a false positive and the second mistake a false negative (with regard to the question as to whether treatment is efficient). I assume throughout that $\alpha + \beta \leq 1$.

In case of remaining uninsured, the consumer only seeks treatment if treatment is deemed worth the cost p (correctly or incorrectly) and if treatment is affordable $x \geq p$. Then the utility from remaining uninsured is given by

$$u_0 = \hat{x} + \pi_h(1 - \rho)(1 - \beta)(V_h - p) - \pi_l(1 - \rho)\alpha(p - V_l). \tag{15}$$

In contrast, full insurance leads to treatment independent of its value:

$$u_{full} = \hat{x} + \pi_h V_h + \pi_l V_l - w. \tag{16}$$

---

applied to insurance rebates and deductibles. The only alternative explanation for such a preference, that I am aware of, is Zweifel (1987). In that framework, rebates are desirable for breaking the time correlation, that is associated with cost-sharing, between a financial loss and a health loss.

[12]Note that this is not true for the affordability constraints that apply to deductibles. Subsidizing deductible payments undermines the incentives that are the reason for imposing cost-sharing in the first place.

Maximal willingness-to-pay and expected cost of insurance are then given by

$$\bar{w}_{full} = \pi_h \left[V_h(1 - (1-\rho)(1-\beta)) + p(1-\rho)(1-\beta)\right] + \pi_l \left[V_l(1-(1-\rho)\alpha) + p(1-\rho)\alpha\right], \quad (17)$$

$$c_{full} = (\pi_h + \pi_l)p. \quad (18)$$

Hence, full insurance creates a surplus of

$$s_{full} - c_{full} = \pi_h \left[\beta + \rho(1-\beta)\right](V_h - p) - \pi_l \left[(1-\alpha) + \rho\alpha\right](p - V_l). \quad (19)$$

As before, s_{full} is not necessarily positive. The first, positive part reflects the consumption of additional efficient care due to full insurance. The second, negative part reflects the consumption of additional inefficient care due to full insurance. It is noteworthy that both mistakes have a positive effect on s_{full}. Consider the implication of a false negative (β): if the consumer wrongly deems high-value care to be of low value, treatment is avoided without insurance. Under full insurance, all care is consumed irrespective of the value that the consumer assigns to it. Thus, the fully-insured consumer receives treatments that are incorrectly deemed of little value. In this way, full insurance corrects the false-negative mistakes of consumers. Consider next the implication of a false positive (α): if the consumer wrongly classifies care as being of high value, it is also consumed by the uninsured (as long as it is affordable). That means that a certain part of inefficient care is consumed regardless of insurance status. If this is true, full insurance is no longer responsible for all of the inefficient care that the consumer demands. This reduces the severity of inefficient moral hazard. Pauly and Blavin (2008) have already pointed out these positive effects of consumer mistakes on the desirability of full insurance. Their analysis, however, misses the dampening impact of affordability constraints on this result. As $\rho \to 1$, consumer mistakes have no effect on the desirability of full insurance anymore. This is intuitive: if access barriers fully prohibit the consumption of care when uninsured, then the beliefs of the consumer do not matter anymore. No care is consumed when uninsured irrespective of perceived value, while all care is consumed by the fully insured irrespective of perceived value. In that case, consumer mistakes neither affect the desirability of remaining uninsured nor the desirability of full insurance, and, hence, they cannot affect the comparison of the two.

Consider again partial insurance with a deductible d that makes the consumption of low-value

care unattractive: $V_l = d < p$. The utility from buying partial insurance is given by

$$u_p = \hat{x} + \pi_h(1 - \delta)(1 - \beta)(V_h - d) + \pi_l(1 - \delta)\alpha(V_l - d). \tag{20}$$

Maximal willingness-to-pay and expected cost of insurance are given by

$$\bar{w}_p = \pi_h(1 - \beta)\left[(1 - \delta)(p - d) + (\rho - \delta)(V_h - p)\right] + \pi_l\alpha\left[(1 - \delta)(p - d) + (\rho - \delta)(V_l - p)\right], \tag{21}$$

$$c_p = (1 - \delta)\left[\pi_h(1 - \beta) + \pi_l\alpha\right](p - d). \tag{22}$$

Hence, cost-sharing through partial insurance creates a surplus of

$$s_p = \bar{w}_p - c_p = \pi_h(1 - \beta)(\rho - \delta)(V_h - p) - \pi_l\alpha(\rho - \delta)(p - V_l). \tag{23}$$

In contrast to the case of the fully-informed decision-maker, this surplus is not necessarily positive. This is because partial insurance can only deter the consumption of inefficient care if it is correctly identified as such, i.e., if the consumer does not commit a false positive. In addition, the incentivization of cost-sharing reduces the consumption of efficient care for two reasons. First, as in the case of the fully-informed consumer, efficient care is not consumed if the deductible is unaffordable. Second, efficient care is no longer consumed even if affordable. If it is wrongly considered as inefficient, i.e., if the consumer commits a false negative, the cost-sharing requirement deters the consumer from seeking high-value care.

Cost sharing has two advantages over full insurance. First, consumption of nonvaluable care is restricted to the case of a false positive. Second, even if a false positive occurs, the individual may no longer consume nonvaluable care whenever the required deductible is unaffordable. Hence, deductible insurance reduces detrimental moral hazard to the case in which both a false positive occurs and the deductible is affordable. These two advantages need to be weighed against two disadvantages. First, consumption of valuable care is restricted to the states in which it is correctly recognized as such. That is, insurance no longer corrects false negatives. In addition, partial insurance confines the consumption of valuable care to the state in which the deductible is affordable and thereby reduces the access value.

Consider the case of rebate insurance if the rebate is set such that low-cost care is not consumed: $r = V_l$. This yields a utility of

$$u_r = \hat{x} + \pi_h(1 - \beta)V_h + \pi_l \alpha V_l + (1 - \pi_h(1 - \beta) - \pi_l \alpha)r - w, \tag{24}$$

which can be used to determine the maximum willingness-to-pay. Again, we can calculate the surplus of rebate insurance by subtracting the cost from the willingness-to-pay.

$$\bar{w}_r = \pi_h(1 - \beta)\left[\rho V_h + (1 - \rho)p\right] + \pi_l \alpha \left[\rho V_l + (1 - \rho)p\right] + (1 - \pi_h(1 - \beta) - \pi_l \alpha)r \tag{25}$$

$$c_r = \pi_h(1 - \beta)p + \pi_l \alpha p + (1 - \pi_h(1 - \beta) - \pi_l \alpha)r \tag{26}$$

$$s_r = \pi_h(1 - \beta)\rho(V_h - p) - \pi_l \alpha \rho(p - V_l) \tag{27}$$

Rebate insurance holds an advantage over partial insurance as affordability constraints no longer restrict the consumption of valuable care when correctly identified. On the other hand, these affordability constraints no longer prevent the consumption of low-value care due to a false positive. If the latter effect dominates the former effect, rebate insurance is inferior to partial insurance. Hence, when consumers make mistakes in assessing the quality of care, cost-sharing can be superior to rebates. As is pointed out in the last section, that cannot be the case when consumers are fully informed.

Given the collection of advantages and disadvantages that the different regimes feature, it is important to describe the regime that maximizes welfare (as measured by surplus) for a given parameter constellation. For this matter, define

$$\phi := \frac{\pi_h(V_h - p)}{\pi_l(p - V_l)}. \tag{28}$$

ϕ measures the expected net efficiency of treatment. If $\phi \geq 1$, then the expected value of care is worth its cost, or, put differently: an uninformed decision-maker would seek care if and only if $\phi \geq 1$.

It turns out that the welfare comparison across insurance regimes is rather straightforward:

Proposition 2. *If $\phi > \frac{1-\alpha}{\beta}$, then the largest welfare (surplus) is generated by full insurance. If $\frac{\alpha}{1-\beta} \leq \phi \leq \frac{1-\alpha}{\beta}$, then the largest welfare (surplus) is generated by rebate insurance. If $\phi < \frac{\alpha}{1-\beta}$, then welfare is maximized by no insurance.*

Proof. First, note that $s_r \geq s_p$ holds if and only if $s_p \geq 0$, with strict inequality for $\delta > 0$. Second, the comparison of s_{full} and s_r yields

$$s_r \geq s_{full} \Leftrightarrow \phi \leq \frac{1-\alpha}{\beta} \tag{29}$$

while $s_r \geq 0$ if and only if $\phi \geq \frac{\alpha}{1-\beta}$.

Finally, $\frac{\alpha}{1-\beta} \leq 1 \leq \frac{1-\alpha}{\beta}$ holds since $\alpha + \beta \leq 1$ by assumption.

$\square$

The intuition for this result is as follows. If ϕ is very large, the expected net benefit of receiving high-value care greatly outweighs the expected net cost of paying for low-value care. Hence, the benefit of correcting a false negative through full insurance outweighs the cost of also treating those who (correctly) deem the treatment not worth the cost. This makes full insurance more desirable than any insurance that relies on consumer incentivization, be it partial or rebate insurance. If ϕ is very low, then the cost of providing access to people who commit a false positive outweighs the benefit of providing access to people who correctly deem treatment valuable. In this case, no insurance is welfare-maximizing. In the intermediate cases, when $\phi \approx 1$, it is optimal to incentivize the consumer through rebate insurance.
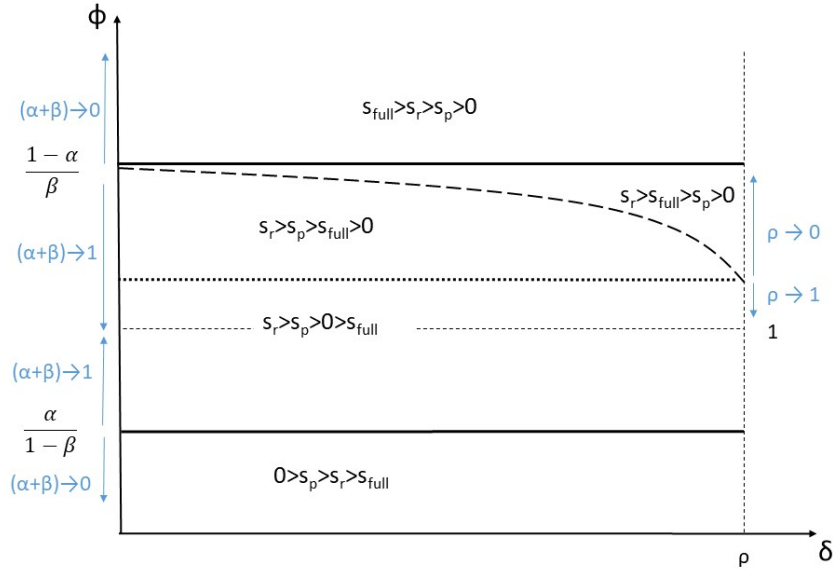

Figure 1: Optimal Insurance Design

Figure 1 illustrates these boundaries. It is only when $\frac{1-\alpha}{\beta} \leq \phi \leq \frac{\alpha}{1-\beta}$ that it is welfare-

maximizing to incentivize the consumer's care decisions through insurance design. Note that this "corridor" spans the entire range as $\alpha + \beta \to 0$. As the consumer's decisions approach the decision of a perfectly-informed decision-maker, rebate insurance is always optimal, as is shown in the previous section. As $\alpha + \beta \to 1$, the range of ϕ, for which incentivization through rebates is optimal, shrinks to zero, because both the upper and the lower boundary converge to 1. In the extreme case, when $\alpha + \beta = 1$, it is either better to fully insure or not to insure at all. This is also intuitive. If $\alpha + \beta = 1$, then the consumer's belief about the quality of care is completely uninformative about the actual quality of care. In that case, it makes no sense to make use of the consumer's "knowledge". Instead welfare maximization follows the decision rule of an uninformed decision-maker who prefers to seek care if and only if $\phi \geq 1$. Hence, if $\phi > 1$, an uninformed decision-maker prefers the consumer to always seek treatment. That is ensured by full insurance. If $\phi < 1$, the decision-maker prefers the consumer to always abstain from treatment, and the treatment probability is minimized if the consumer remains uninsured.

Figure 1 also shows that if there are major affordability barriers imposed by cost-sharing, $\delta \approx \rho \approx 1$, full insurance dominates cost-sharing for almost all values $1 \leq \phi \leq \frac{1-\alpha}{\beta}$. This is because, as $\delta \to 1$, the utilization reduction in response to cost-sharing reveals more information about the consumer's budget than about the consumer's belief about the quality of care. There exists a cutoff $\phi' = \frac{1-(1-\delta)\alpha}{1-(1-\delta)(1-\beta)}$, depicted by the dashed line, above which it is better to fully insure instead of relying on the contaminated information revealed through cost-sharing. In the same way, if $\delta \approx \rho \approx 1$, the surplus generated by cost-sharing in the range $\frac{\alpha}{1-\beta} \geq \phi \geq 1$ vanishes. In conclusion, affordability constraints greatly diminish the informational value provided by cost-sharing. This is not the case for rebate insurance, which is able to fully extract the informational advantage of the consumer - provided there is one ($\alpha + \beta < 1$).

Proposition 2 shows that cost-sharing is never the welfare-maximizing design as long as $\delta > 0$, despite the previous observation that there are cases in which it dominates rebate insurance. However, these turn out to be equivalent to the cases in which no insurance is welfare-maximizing. The intuition of this result is straightforward. In comparison to rebates, cost-sharing has the advantage of restricting the consumption of inefficient care if the consumer commits a false-positive but does not have enough resources to pay for the deductible. This advantage has to be weighed against the disadvantage of consumers not being able to receive (correctly-identified) efficient care when deductibles are unaffordable. If the advantage of cost-sharing is larger than the disadvantage, cost-sharing is better than rebates in terms of welfare. However, if eliminating the (erroneous)

consumption of inefficient care under a false positive is worth restricting the consumption of (correctly identified) efficient care, then this is true regardless of a person's budget. Hence, it also applies to the case in which the deductible is affordable. Importantly, the same advantage and disadvantage - restricted to the case when the deductible is affordable - describes the comparison of partial insurance to remaining uninsured. Thus, if (and only if) cost-sharing dominates rebates, then no insurance dominates cost-sharing.

While section 2 shows that rebate insurance is maximizing welfare if the consumer is perfectly informed, this section shows that this result still holds if the consumer is informed sufficiently well. As the consumer's information deteriorates, the extreme options of full or no insurance become the designs that maximize welfare. Apart from boundary cases, $\delta = 0$, cost-sharing is never the optimal solution. This is because cost-sharing basically trades one inefficiency for another as it reduces the consumption of efficient care along with the intended reduction of inefficient care. As it turns out, the inefficiency of reduced access through cost-sharing breeds an additional inefficiency: adverse selection.

4   Adverse Selection on Income

Adverse selection has typically been described with reference to heterogeneity in the health risk π. Here, I want to point out that cost-sharing produces an additional source of adverse selection. Assume a perfectly-informed consumer and consider the expected cost of insurance under cost-sharing

$$c_p = \pi_h(1 - \delta)(p - d). \tag{30}$$

Beyond the health risk π_h, the cost of insurance is also determined by the probability of an insuree being able to pay the deductible $(1 - \delta)$. If people differ in their ability to pay, then there is cost heterogeneity even after controlling for health risk. Simply put, the "height" of access barriers imposed by cost-sharing may differ across people depending on their financial status. Formally, suppose people are described by their type of budget risk $\theta \in \{H, L\}$, with a high risk facing a larger probability of not being able to afford a payment of size x: $F_H(x) > F_L(x)$, $\forall x > 0$. Straightforwardly, this implies $c_p(H) = \pi_h(1 - \delta_H)(p - d) < \pi_h(1 - \delta_L)(p - d) = c_p(L)$, where $\delta_\theta = F_\theta(d)$. That means it is more costly to insure the more affluent, i.e. those who face a

lower budget risk. Note that, at the same time, the first-best surplus from insurance is given by $s^{FB}(\theta) = \pi_h \rho_\theta (V_h - p)$ with $\rho_\theta = F_\theta(p)$. The affluent, i.e., the type with the lower budget risk, derives a lower net value from insuring.

Given that the poor are less costly to insure, they face selective pressure. Either the poor are pooled with the rich and are, thereby, forced to cross-subsidize the more extensive utilization of the affluent while themselves facing a major probability of being unable to actually use their health plan once in need. Or they can choose to leave the market foregoing any gains from insuring. Cost-sharing thus creates a selective pressure against the very type that derives the largest value from insurance.

The selective pressure on the poor is not present under rebate insurance for the simple reason that they do not impose any access barriers on using a given plan. Hence, conditional on health risk, both types have identical cost of insurance

$$c_r(H) = (1 - \pi_h)r + \pi_h p = c_r(L). \tag{31}$$

Consequently, there is no selective pressure created by pooling the two types, resulting in the first-best surplus for both types.

By creating access barriers, cost-sharing actually produces two novel inefficiencies in the attempt to combat inefficient moral hazard. First, it restricts the consumption of efficient care by imposing affordability constraints on the consumption of said care. Second, as these constraints differ across income groups, they produce a source of adverse selection that puts selective pressure on exactly the type that, conditional on health status, derives the largest value from insurance.

Scholars have already recognized that income differences can be a source of heterogeneity with regard to insurance cost and/or value, and have analyzed the implications for market outcomes. Wambach (2000) analyzes an insurance market in which types differ with regard to their health risk and their wealth. The latter determines insurance value - but not its cost - by assuming decreasing risk aversion. Boone and Schottmüller (2015) show how income differences may explain the phenomenon of advantageous selection if income is negatively correlated with health risk and positively correlated with care utilization in case of sickness.[13] Here, I argue that an alternative

---

[13]Boone and Schottmüller (2015) assume utilization to vary on the intensive margin with income, while access problems are more in line with variation on the extensive margin. Still, Boone and Schottmüller (2015) is the only work, that I am aware of, that recognizes the potential of cost-sharing to be a source of adverse selection.

incentivization scheme based on boni can mitigate this source of heterogeneity.

It is noteworthy that rebates have so far been typically analyzed as a means to screen different risk types.[14] This literature analyzes the usefulness of rebates as a tool for risk-adjustment in a dynamic setting. The literature has thus already established the potential usefulness of such an incentivization to address adverse selection.

5   Conclusion

Cost-sharing is observed to reduce care utilization, but the observed reduction does not seem to distinguish sufficiently with respect to the quality of care. In this paper, I use a simple model that incorporates two potential reasons for this indiscriminate response: limited access and consumer mistakes. The analysis reveals that cost-sharing is never a welfare-maximizing insurance design. If consumers are sufficiently well-informed about the quality of care, then it is desirable to extract that information through appropriate incentivization. In this case, however, a positive incentivization through bonus payments/rebates dominates a negative incentivization through cost-sharing, as the former does not restrict consumption of efficient care based on affordability. It is shown that, alongside deterring the consumption of inefficient care, cost-sharing produces two new inefficiencies by restricting access to efficient care and, thereby, producing an adverse selection problem based on income. Rebate insurance yields the same incentivization without producing these inefficiencies. If consumer mistakes are strong enough, the extremes of full or no insurance become the welfare-maximizing regime, as there is little use in incentivizing an ill-informed consumer.

The paper shows that limitations of access are no necessary byproduct of incentivizing the efficient utilization of care. Depending on the relative importance of efficient and inefficient care, cost-sharing may actually do more harm than good. In this paper, a simple, alternative design invoking bonuses is proposed that is already in use in some insurance markets - although mostly for the purpose of risk adjustment. Offering positive incentives to consumers for not using inefficient care can yield the same incentives as negative incentives. At the same time, they allow consumers to express their preferences, instead of just revealing their financial capabilities. This points at a more general lesson. If affordability problems distort the appropriate reaction to incentives, then it can be welfare-enhancing to use shadow prices in a positive incentive scheme instead of using

---

[14]See Dionne, Fombaron, and Doherty (2013) for a nice exposition.

actual prices in a negative incentive scheme.

References

Beck, R. (1974). The effects of co-payment on the poor. *The Journal of Human Resources 9 (1)*, 129–142.

Boone, J. and C. Schottmüller (2015). Health insurance without single crossing: Why healthy people have high coverage. *The Economic Journal 127*, 84–105.

Brot-Goldberg, Z., A. Chandra, B. Handel, and J. Kolstad (2015). What do deductibles do? The impact of cost-sharing on health care prices, quantities, and spending dynamics. *NBER Working Paper Series, No. 21632*.

Chandra, A., J. Gruber, and R. McKnight (2014). The impact of patient cost-sharing on low-income populations: Evidence from Massachusetts. *Journal of Health Economics 33*, 57–66.

Dionne, G., N. Fombaron, and N. Doherty (2013). Adverse selection in insurance contracting. In G. Dionne (Ed.), *Handbook of Insurance*. Springer.

Fels, M. (2016). Mental accounting, access motives, and overinsurance. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2611896`.

Fendrick, A. and M. Chernew (2006). Value-based insurance design: Aligning incentives to bridge the divide between quality improvement and cost containment. *The American Journal of Managed Care 12*, SP5–SP10.

Fendrick, A., D. Smith, M. Chernew, and S. Shah (2001). A benefit-based copay for prescription drugs: Patient contribution based on total benefits, not drug acquisition cost. *The American Journal of Managed Care 7*(9), 861–867.

Jan, T. (2015). Critics say high deductibles make insurance 'unaffordable'. *The Boston Globe*.

Johnson, E., J. Hershey, J. Meszaros, and H. Kunreuther (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty 7 (1)*, 35–51.

Kunreuther, H., M. Pauly, and S. McMorrow (2013). *Insurance & Behavioral Economics: Improving Decisions in the Most Misunderstood Industry*. Cambridge University Press.

McGuire, T. (2011). Demand for health insurance. In M. Pauly, T. McGuire, and P. P. Barros (Eds.), *Handbook of Health Economics, Vol. 2*. Elsevier Science B.V.

Nyman, J. (1999a). The economics of moral hazard revisited. *Journal of Health Economics 18*, 811–824.

Nyman, J. (1999b). The value of health insurance: the access motive. *Journal of Health Economics 18*, 141–152.

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

Pear, R. (2015). Many say high deductibles make their health law insurance all but useless. *The New York Times*.

Peele, P. (1993). Evaluating welfare losses in the health care market. *Journal of Health Economics 12*, 205–208.

Rice, T. (1992). An alternative framework for evaluating welfare losses in the health care market. *Journal of Health Economics 11*, 85–92.

Rubinstein, A. and M. E. Yaari (1983). Repeated insurance contracts and moral hazard. *Journal of Economic Theory 30*, 74–97.

Schokkaert, E., J. Steel, and C. Van de Voorde (2017). Out-of-pocket payments and subjective unmet need of healthcare. *Applied Health Economics and Health Policy*, 1–11.

Tamblyn, R., R. Laprise, J. Hanley, M. Abrahamowicz, S. Scott, N. Mayo, J. Hurley, R. Grad, E. Latimer, R. Perreault, P. McLeod, A. Huang, P. Larochelle, and L. Mallet (2001). Adverse events associated with prescription drug cost-sharing among poor and elderly persons. *Journal of the American Medical Association 285*, 421–429.

Wambach, A. (2000). Introducing heterogeneity in the Rothschild-Stiglitz model. *The Journal of Risk and Insurance 67 (4)*, 579–591.

Zweifel, P. (1987). Bonus systems in health insurance: a microeconomic analysis. *Health Policy 7*, 273–288.

Zweifel, P. and W. Manning (2000). Moral hazard and consumer incentives in health care. In A. Culyer and J. Newhouse (Eds.), *Handbook of Health Economics, Vol. 1*. Elsevier Science B.V.

Working Paper Series in Economics

recent issues