Received: 17 September 2019 | Revised: 20 February 2020 | Accepted: 6 April 2020

DOI: 10.1002/hec.4023

RESEARCH ARTICLE

WILEY Health Economics

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

Markus Fels

Department of Economics, University of Dortmund (TU), Dortmund, Germany

Correspondence
Markus Fels, University of Dortmund (TU), Department of Economics, Vogelpothsweg 87, 44227 Dortmund, Germany.
Email: markus.fels@udo.edu

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. I show how the optimal design can be deduced empirically and discuss possible impediments to its implementation.

KEYWORDS

cost-sharing, insurance rebates, limited access, moral hazard

JEL CLASSIFICATION

D82; I13; I14

1 | INTRODUCTION

Moral hazard, specifically ex post moral hazard, is argued to be one of the main impediments to a well-functioning insurance market. [1] 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 cost-sharing as a means to deter the consumption of inefficient care. 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.[2] Reducing these additional

[1] For nice overviews of the literature on moral hazard in health insurance, see Zweifel and Manning (2000) and McGuire (2011).
[2] Nyman (1999b) calls this benefit of insurance its access value.

expenditures through cost-sharing is then neither necessary nor desirable.[3]

Empirical evidence is robust in showing that cost-sharing is successful in reducing utilization (Zweifel & 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 & Manning, 2000; Brot-Goldberg, Chandra, Handel, & Kolstad, 2017). 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).[4] Pauly and Blavin (2008) suggest to differentiate cost-sharing arrangements according to consumers' ability in appropriately assessing the quality of care. This idea is essentially reflected in the concept of value-based insurance (Fendrick, Smith, Chernew, & Shah, 2001; Fendrick & Chernew, 2006). Still, researchers remain worried that cost-sharing arrangements are an instrument that discriminates only insufficiently between appropriate and inappropriate care (Zweifel & 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.[5] This criticism is supported by evidence that suggests a negative association between insurance coverage and health outcomes for those with low income and poor health (Zweifel & 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.

In this paper, I analyze different insurance designs to address moral hazard.[6] 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 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.[7] In the following section, I extend the model to allow for consumer mistakes in distinguishing low-value from high-value care. I then 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. Again, I find cost-sharing to never be an optimal design. In the following section, I discuss several issues arising from the original analysis. First, I point out that cost-sharing leads to an additional inefficiency. 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. Second, I propose a strategy on how to disentangle empirically the two channels of the utilization response to cost-sharing and show how the welfare-maximizing insurance design can be deduced from utilization data. Finally, I discuss barriers to implementing the efficient design in a private market.

2 | AN INSURANCE MODEL WITH TWO TYPES OF MORAL HAZARD

Suppose an 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 either confer a high value $V_h > p$, such that treatment is efficient, or it can be of low value $V_l < p$, such that it is efficient to abstain from treatment.

[3]In fact, Fels (2020) shows that deductibles can destroy more than the access value of insurance if affordability constraints matter. In addition, they 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.

[4]Others defend the traditional welfare analysis, insisting that treatment effectiveness is not to be confused with treatment efficiency (Peele, 1993).

[5]See Beck (1974) for early evidence of a stronger response to cost-sharing among lower-income households. In contrast, Chandra, Gruber, and McKnight (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, that is, a reduction of utilization to zero. Schoen et al. (2010) survey access impediments in eleven countries. For a recent review of the literature on the relationship between out-of-pocket cost and utilization, see Schokkaert, Steel, and Van de Voorde (2017).

[6]Boone (2018) provides a nice analysis which conditions should be covered by basic or supplementary insurance if patients face access problems. Here, I focus on the optimal design of incentives. Besanko, Dranove, and Garthwaite (2016) analyze how coverage by insurance that offers an access value influences a monopolist's decision to price pharmaceuticals.

[7]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.

Ex ante, π_h denotes the joint probability that an individual becomes sick and treatment has high value, and π_l denotes the joint probability that an individual becomes sick and treatment has low value, 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 denote the cumulative distribution function over X, and by $\hat{X} = \int xdF(x)$ the expected wealth at the time of need. Assume a simple additively separable utility function $u(c, k) = c + k$ where c denotes nonmedical consumption and k denotes utility from receiving care.[8] Without loss of generality, 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 expected utility of an uninsured individual 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 $\rho = F(p) = Prob(X < p)$ denotes the probability of not being able to afford the cost of care p. An uninsured individual who has to pay the full cost of care out of pocket decides to receive treatment only if it is of high value. Yet, even if treatment has high value, the consumer may not be able to afford treatment when his budget falls below p.

Suppose that the individual has insurance that fully covers treatment cost. As the price that the consumer needs to pay to receive treatment is reduced to zero, care is consumed irrespective of value. An individual who is fully insured receives an expected utility of

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

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. The value of full insurance, calculated as the utility difference $\bar{w}_{full} = u_{full} - u_0$, is then given by

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

The expected cost of associated with providing the individual with full insurance is given by $c_{full} = \pi p$. This allows to derive the 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 surplus is a useful measure to derive the welfare consequences of insurance. It is identical to the gains from trade in an insurance market in which the individual is willing and able to buy full coverage and an insurer is able to provide full coverage at fair cost c_{full}. Equally, it measures the welfare gain from a public insurance scheme in which full coverage can be financed without any efficiency loss from taxation. The surplus thus offers an abstract measure of the net value of insurance that is independent of the institutional arrangement of how the cost c_{full} is covered. This independence allows to separate the analysis of the insurance regime from the discussion which institutional arrangement is able to implement the regime and at what cost.[9]

The surplus of full insurance consists of two parts. The first part reflects the access value of insurance as described by Nyman (1999b). Insurance is valuable (even to an individual who is not risk averse) 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 receive 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 result in a negative surplus, leaving no gains from trade in a private market or

[8]The assumption of utility being linear in c rules out an insurance motive based on risk aversion. This is a deliberate modeling choice to simplify the exposition. It departs from the conventional assumption that risk aversion is the sole motive for insurance purchase. Instead, it follows Nyman's suggestion that the major purpose of health insurance is the access motive (Nyman, 1999b). The analysis in Fels (2019) shows that the access value indeed dominates any consumption-smoothing value of insurance for insurances like health insurance that cover imperfectly divisible goods or services that have value only in a subset of states.

[9]See the discussion in Section 4.4.

resulting in a welfare decrease through public insurance. Accordingly, economists have proposed to impose cost-sharing as a means to deter the consumption of ineffective care, thereby eliminating inefficient moral hazard.

Suppose that insurance does not fully cover 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 seeks care if it is of high value. An individual who has insurance with such a cost-sharing requirement then receives an expected utility of

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

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

where $\delta = F(d) = Prob(X < d)$. The value and the cost of insurance with cost-sharing are given by

$$\bar{w}_{cs} = u_{cs} - u_0 = \pi_h \left[(1 - \delta)(p - d) + (\rho - \delta)(V_h - p)\right], \tag{8}$$

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

This results in a surplus of

$$s_{cs} = w_{cs} - c_{cs} = \pi_h(\rho - \delta)(V_h - p). \tag{10}$$

The nonnegativity of the surplus of partial insurance shows that cost-sharing is 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, cost-sharing is 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, cost-sharing can amount to throwing out the baby with the bathwater.

Suppose that, instead of requiring a deductible payment in case of the consumption of care, insurance fully covers treatment costs, 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

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

Calculating the value, the expected cost, and the surplus yields

$$\bar{w}_r = u_r - u_0 = (1 - \pi_h)r + \pi_h(p + \rho(V_h - p)), \tag{12}$$

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

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

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 the 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 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, $\delta > 0$.

[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 refer to the total amount as the deductible.

Proposition 1. *If consumers are perfectly informed, rebate insurance is able to fully eliminate inefficient moral hazard while protecting access, thereby maximizing the surplus from insurance (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[11] can straightforwardly explain evidence that suggests a preference for rebates over deductibles as observed by Johnson, Hershey, Meszaros, and Kunreuther (1993).[12] It is, however, important to recognize that rebate insurance requires premium payments that are substantially larger than insurance relying on cost-sharing. Hence, affordability constraints are likely relevant at the time of insurance purchase.[13]

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 & Blavin, 2008). Consumers may reduce both the consumption of effective and ineffective care in response to cost-sharing if they have problems distinguishing between 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 barriers 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 of 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 that $\alpha + \beta \leq 1$.

Without insurance, 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 expected utility of the 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 and thus yields an expected utility of

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

The value and the expected cost of insurance are then given by

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

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

Hence, full insurance creates a surplus of

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

As before, s_{full} is not necessarily positive. The first, positive part reflects the consumption of additional efficient care and 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

[11] The welfare gains from rebates in comparison to cost-sharing directly relate to Nyman's access value of insurance. Nyman (2003, p. 102 ff.) shows how the access value of insurance can be estimated. The welfare gains from rebates can be derived by multiplying an estimate of the access value with an estimate of the probability δ that an individual is unable to afford the cost-sharing requirement.

[12] See also Kunreuther, Pauly, and McMorrow (2013, pp. 118–119). Johnson et al. (1993) propose an explanation based on different frames being applied to insurance rebates and deductibles. The only alternative explanation for such a preference, which I am aware of, is Zweifel (1987). In that framework, rebates are desirable for breaking the time correlation between a financial loss and a health loss.

[13] See the discussion in Section 4.4.

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 low-value care is wrongly classified as having high value, it is consumed also 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 these constraints tighten, $F(p) = \rho \to 1$, 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, whereas all care is consumed by the fully insured irrespective of perceived value. In that case, consumer mistakes do not affect the value of full insurance.

Consider again partial insurance with a deductible d that makes the consumption of low-value care unattractive: $V_l = d < p$. The utility of an individual with partial insurance is then given by

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

Maximum willingness to pay and expected cost of insurance are given by

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

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

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

$$s_{cs} = w_{cs} - c_{cs} = \pi_h(1-\beta)(\rho-\delta)(V_h - p) - \pi_l\alpha(\rho-\delta)(p - V_l). \tag{23}$$

In contrast to the case of the fully informed decision-maker, this surplus is not necessarily positive. This is because cost-sharing can only deter the consumption of inefficient care if it is correctly identified as such, that is, 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 inefficient, that is, 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 low-value care is restricted to the case of a false positive. Second, even if a false positive occurs, the individual may not consume low-value care if the required deductible is unaffordable. Hence, cost-sharing reduces inefficient 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 high-value care is restricted to the states in which it is correctly recognized as such. That is, insurance no longer corrects false negatives. In addition, cost-sharing 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-value care is not consumed: $r = V_l$. This yields a utility of

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

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

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

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

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

Rebate insurance holds an advantage over cost-sharing 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 previous 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 seeks care if and only if $\phi \geq 1$.

It turns out that the welfare comparison across insurance regimes is rather straightforward. Define $\bar{\phi} := \frac{1-\alpha}{\beta}$ and $\underline{\phi} := \frac{\alpha}{1-\beta}$.

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

First, note that $s_r \geq s_{cs}$ holds if and only if $s_{cs} \geq 0$, with strict inequality for $\delta > 0$.

Second, the comparison of s_{full} and s_r yields

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

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

Finally, $\underline{\phi} \leq 1 \leq \bar{\phi}$ holds because $\alpha + \beta \leq 1$ by assumption.

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 a treatment not worth the cost. This makes full insurance more desirable than any insurance that relies on consumer incentivization, be it through cost-sharing or rebates. 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 illustrates these boundaries. It is only when $\underline{\phi} \leq \phi \leq \bar{\phi}$ that it is welfare-maximizing to incentivize the consumer's utilization decision through insurance design. Note that this "corridor" spans the entire range of ϕ for a perfectly informed consumer as $\alpha + \beta \to 0$ implies $\underline{\phi} \to 0$ and $\bar{\phi} \to \infty$. As the consumer's decisions get closer to the decisions of a perfectly informed decision-maker, rebate insurance is optimal for almost all values of ϕ, as is shown in the previous section. In contrast, as $\alpha + \beta \to 1$, the range of ϕ, for which incentivization through rebates is optimal, shrinks to zero, because both $\underline{\phi}$ and $\bar{\phi}$ 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 a cutoff $\tilde{\phi} = \frac{1-(1-\delta)\alpha}{1-(1-\delta)(1-\beta)}$, 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 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 exactly the cases in which no insurance is welfare-maximizing. The intuition of this result is straightforward. In comparison to rebates,

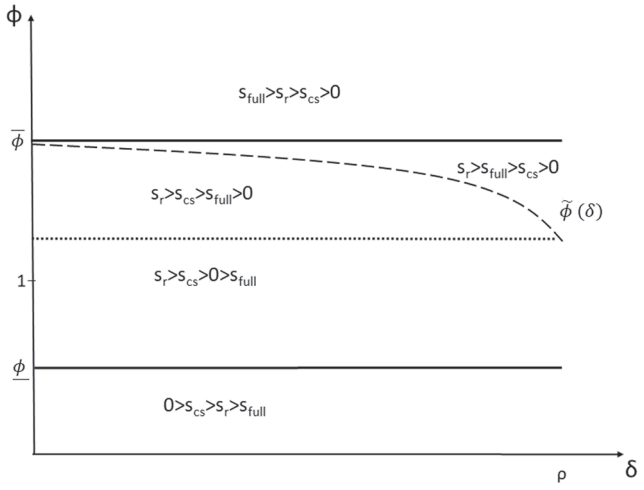

FIGURE 1 Optimal insurance design

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 cost-sharing to remaining uninsured. Thus, if (and only if) cost-sharing dominates rebates, then no insurance dominates cost-sharing.

Whereas 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.

4 | DISCUSSION

4.1 | Adverse selection on income

The analysis from the simple model entirely abstracts from any heterogeneity between individuals. Still, it can be shown that cost-sharing leads to an inefficiency by restricting access to care. Here, I want to shortly point to another potential source of inefficiency from cost-sharing based on income differences.

Adverse selection has typically been described with reference to heterogeneity in the health risk π. With budget constraints inhibiting access, cost-sharing produces an additional source of adverse selection. Assume a perfectly informed consumer and consider the expected cost of insurance under cost-sharing

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

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 z: $F_H(z) > F_L(z), \forall z > 0$. Straightforwardly, this implies $c_{cs}(H) = \pi_h(1-\delta_H)(p-d) < \pi_h(1-\delta_L)(p-d) = c_{cs}(L)$, where $\delta_\theta = F_\theta(d)$. It is more costly to insure the more affluent, that is, those who face a lower budget risk, as they are more likely to seek treatment given sickness. In contrast, the first-best surplus from insurance is given by $s^{FB}(\theta) = \pi_h\rho_\theta(V_h - p)$ with $\rho_\theta = F_\theta(p)$. Hence, the more affluent, that is, the type with the lower budget risk, derives a lower net value from insuring.

Suppose insurance premiums do not discriminate according to income. 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 reduce 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.[14] Here, I argue that an alternative incentivization scheme based on bonuses 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.[15] 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 a design to address adverse selection on risk types. This section points out an additional use by preventing adverse selection on income.

4.2 | Assessing the importance of the two channels empirically

Although cost-sharing is shown to never be an optimal insurance regime if it produces access barriers, the optimal insurance regime still depends on the importance of consumer mistakes in assessing treatment value. In particular, the optimality of using any consumer incentivization strongly depends on the size of choice errors. In addition, the inferiority of cost-sharing to rebates depends on the (relative) importance of access barriers for explaining the indiscriminate response of utilization to cost-sharing requirements. It is thus important to assess the relative importance of the two potential channels—access barriers and consumer mistakes—empirically.

Denote by $\mu_k, k \in \{f, cs, r\}$ the utilization of health services that is observed under full insurance (f), cost-sharing (cs), and rebate insurance (r).

$$\begin{aligned}\mu_f &= [\pi_h + \pi_l]p,\\ \mu_{cs} &= [\pi_h(1-\delta)(1-\beta) + \pi_l(1-\delta)\alpha]p,\\ \mu_r &= [\pi_h(1-\beta) + \pi_l\alpha]p.\end{aligned}$$

The model then predicts the following association between the relative reduction in utilization in response to cost-sharing, $\Delta_{cs} := ((\mu_f - \mu_{cs})/\mu_f)$, and the relative reduction in utilization in response to rebates $\Delta_r := ((\mu_f - \mu_r)/\mu_f)$:

$$\Delta_{cs} = \Delta_r + \delta\frac{\mu_r}{\mu_f}. \tag{32}$$

The first term on the right-hand side captures the part of the reduction in utilization that can be attributed to the incentives that both cost-sharing and rebates provide. The second term captures the part of the utilization response that is

[14] 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, which I am aware of, that recognizes the potential of cost-sharing to be a source of adverse selection.

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

associated with access barriers and thus unique to cost-sharing. If we have data on the utilization response to cost-sharing and the utility response to rebates, we can thus disentangle the two effects. For example, if we observe that consumers do not respond to rebates at all ($\Delta_r = 0$), we must conclude that the entire response to cost-sharing is due to access barriers. On the other hand, a similar response to cost-sharing and rebates indicates a low relevance of access barriers in the utilization decision.

Note that this implicitly assumes that the only reason for a dissimilar response are access barriers. If people respond stronger to cost-sharing for other reasons, such as loss aversion, this overestimates the significance of access barriers.[16] One way to control for that is to check whether the difference between cost-sharing and rebates is more pronounced for low-income groups as opposed to high-income groups. Hayen, Klein, and Salm (2018) apply this procedure to data from the Netherlands and find that access barriers played no role in the utilization response to a cost-sharing requirement in the Dutch health insurance system.

Hence, data on the utilization response to cost-sharing and to rebates can help us assess the importance of access barriers. If there is a difference in response, then access barriers matter, and thus rebates dominate cost-sharing as long as insurance is desirable.

Another nice example is provided by Remmerswaal, Boone, Bijlsma, and Douven (2019), who found no significant response in the utilization of lower-income households to rebates but a significant reduction in utilization in response to deductibles. In stark contrast, they find higher-income households to respond to both types of incentivization with no significant difference in the reduction of utilization due to rebates and due to cost-sharing. The model allows to draw the following conclusions from this data.

First, the response of lower-income households to cost-sharing seems to be entirely based on access barriers and not driven by incentives. Hence, an optimal insurance design for lower-income households should not rely on incentives (neither through rebates nor through cost-sharing) as the reaction of those households seems uninformative of actual treatment value. Instead, the optimal insurance fully covers any expenses that have a sufficiently high expected value and offers no coverage for treatments with low expected value.

Second, the similar response of higher-income households to rebates and deductibles reveal that access barriers are of little importance in the utilization response of these households. Hence, the utilization response is entirely based on incentives, that is, that the utilization response is driven by the household's belief about treatment quality. What we cannot conclude from the data is whether the treatment choices are sufficiently well-informed to justify an incentivization, for it does not reveal information about the precision of the household's beliefs. This is the topic that I turn to now.

4.3 | Assessing the welfare implications of insurance empirically

Few studies can directly compare the utilization response to both cost-sharing and rebates as we typically see only one of the regimes implemented. Still, the model is useful in deriving the welfare implications of the implemented design in comparison to full and no insurance.

Suppose that it is possible for the researcher to distinguish high-value from low-value care, for example, because expert opinions allow for an informed guess. The utilization response can then be further decomposed according to the the value of care. Let a superscript $\theta \in \{h, l\}$ specify the utilization response for high- (h) and low-value (l) care. Then

$$\Delta_{cs}^{h} = \Delta_{r}^{h} + \delta \frac{\mu_r^h}{\mu_f^h} = \beta + \delta(1-\beta),$$

$$\Delta_{cs}^{l} = \Delta_{r}^{l} + \delta \frac{\mu_r^l}{\mu_f^l} = (1-\alpha) + \delta\alpha.$$

[16]This classic prediction by Johnson et al. (1993) was tested in a recent study by Fels, Kairies-Schwarz, and Waibel (2020), who found no evidence of a stronger response to cost-sharing due to loss aversion. On the contrary, when the only difference between rebates and cost-sharing is due to framing, the utilization response to rebates is significantly stronger. The authors argue that this effect is due to mental accounting and a lower salience of dynamic incentives under the rebate frame. The stronger response to rebates due to framing suggests that the discussed procedure actually provides a very conservative estimate of the existence of access barriers.

Note that, for $\Delta_{cs}^{l} < 1$, we can deduce the lower bound $\underline{\phi}$:

$$\frac{1-\Delta_{cs}^{l}}{1-\Delta_{cs}^{h}} = \frac{\alpha}{1-\beta} =: \underline{\phi}. \tag{33}$$

In addition, for $\Delta_{cs}^{l} > 0$, we can deduce the bound $\tilde{\phi}$:

$$\frac{\Delta_{cs}^{l}}{\Delta_{cs}^{h}} = \frac{1-(1-\delta)\alpha}{1-(1-\delta)(1-\beta)} =: \phi. \tag{34}$$

Hence, with data on the response to cost-sharing alone, we can already deduce $\underline{\phi}$ and $\tilde{\phi}$. Given some hunch on ϕ, this allows for welfare comparisons of cost-sharing to full insurance and no insurance.

As an illustration, consider the reduction in hospital days in response to cost-sharing as reported in Siu et al. (1986). In response to cost-sharing, inappropriate days fell by roughly 27.1% and appropriate days fell by roughly 20.6%. Taking these as rough indicators for the decline in appropriate and inappropriate spending on hospital care in response to cost-sharing leads to estimates of $\underline{\phi} \approx 0.92$ and of $\tilde{\phi} \approx 1.31$. Hence, an incentivization of the consumption of hospital days through cost-sharing makes sense only if the expected net value of an appropriate day is between 92% and 131% of the expected net cost of an inappropriate day.[17]

Similarly, suppose we only have data on the utilization response to rebates. For $0 < \Delta_{r}^{h} < 1$, we can then deduce the lower and upper bounds that are necessary to evaluate rebates:

$$\frac{1-\Delta_{r}^{l}}{1-\Delta_{r}^{h}} = \frac{\alpha}{1-\beta} =: \underline{\phi}, \tag{35}$$

$$\frac{\Delta_{r}^{l}}{\Delta_{r}^{h}} = \frac{1-\alpha}{\beta} =: \bar{\phi}. \tag{36}$$

Again, we can deduce the boundaries $\underline{\phi}, \bar{\phi}$, which determine whether rebate insurance dominates full insurance and no insurance, from utilization data.

4.4 | Implementing the optimal design

So far, I have analyzed the welfare implications of different insurance designs, specifically the optimal design of incentive schemes in insurance. Furthermore, I have shown how to determine the optimal regime based on empirical data. Once the optimal regime is determined, it needs to be implemented in a private or public insurance scheme. This section seeks to present some considerations with regard to the implementation. Implementation is, of course, only warranted in cases in which some insurance is desirable $\phi > \underline{\phi}$. This leaves two cases. First, the case in which incentivization of utilization is desirable through a system of rebates, $\phi \in \left[\underline{\phi}, \bar{\phi}\right]$, and, second, the case in which it is undesirable because of sufficiently severe mistakes, and individuals are best off with a system of full insurance, $\phi \geq \bar{\phi}$.

Here, I want to point out that it can be impossible to implement the efficient design in a purely private insurance market. This may be obvious in the case of full insurance, as people with low financial means will have difficulties even paying the fair premium of insuring all efficient care $\pi_h p$. Much less will they have the means to pay the fair premium for an insurance covering all care, efficient or inefficient.

However, a similar problem occurs in a system that seeks to incentivize against the utilization of inefficient care through a rebate system. For the surplus s_r to realize in a purely private market, an individual must be able to afford to pay at least the fair premium c_r for such an insurance. Note, however, that with $r = d = V_l$, it holds that $c_r \geq c_{cs} + d$. That is, an individual who is able to afford c_r must also be able to pay the premium of an insurance contract with cost-sharing requirement and retain an amount that is sufficient to pay the deductible d if a need occurs. However, empirical evidence suggests that cost-sharing requirements do impose access barriers. Such evidence then either suggests there are no true affordability constraints with respect to cost-sharing requirements and access barriers are a simple consequence of either

[17] I do not intend to make a claim whether and to what extend hospital stays should be covered by health insurance. I only seek to illustrate how the model may be used to inform such decisions based on empirical data.

limited liquidity and/or limited commitment. Or it suggests that there are true affordability constraints with respect to cost-sharing requirements, in which case a rebate system in itself cannot offer a solution as it only moves the affordability problem from the insurance utilization to the insurance purchase stage. This leads to a simple insight: *if affordability constraints limit utilization under cost-sharing, a purely private market without any form of redistribution cannot implement the efficient regime*. Some form of redistribution is necessary to increase efficiency.

This begs the question whether one cannot address the problem of affordability of cost-sharing directly through some form of redistribution. This would avoid channeling more money through the insurance system via rebates. Three possibilities present themselves: subsidizing cost-sharing payments directly, subsidizing insurance premiums, or providing a direct transfer that increases the spending ability of those with limited means. The subsidization of cost-sharing would certainly decrease the associated access barriers. However, it would also undermine the incentives created by cost-sharing, the very reason for which cost-sharing is proposed in the first place. Similarly, subsidizing the premiums of cost-sharing plans could lead consumers to seek more generous plans, again undermining incentives. Generally, cost-sharing creates an inverse relationship between the strength of incentives and the level of plan generosity (Cutler & Zeckhauser, 2000; Jaspersen & Richter, 2015). Hence, subsidies distort incentives either directly or indirectly when plans impose cost-sharing. This is not true for rebate insurance. Here, the more generous plan offers larger bonuses and, thus, there is a positive relationship between the strength of incentives and the level of plan generosity. Subsidizing the premium of rebate insurance thus increases demand for plans with stronger incentives on utilization.

Finally, one could consider employing means-tested transfers that ensure that individuals with insufficient means can afford cost-sharing requirements should the need arise. Formally, this leads to an equivalence between the cost-sharing and rebate regime. Still, rebate insurance might have an important advantage in practice. In order to ensure that the utilization response to cost-sharing is solely based on incentives, one needs to assure that the transfer d, that is made at the beginning of an accounting period, is still available when a need arises during the accounting period. It might turn out rather demanding for a low-income household facing a wide range of financial pressures to keep a specified amount untapped for the a-priori unlikely event of severe sickness. In contrast, rebate insurance locks in the amount $d = r$ over the accounting period, thereby avoiding the possibility that this transfer d is used for a different purpose before the accounting period ends. In this way, rebates may serve as a commitment device. An alternative commitment device could be a transfer into a health savings account (HSA). Such accounts have been introduced in the United States in part to tackle the access barriers associated with high-deductible plans.[18] A transfer into an HSA presents a policy-maker with a dilemma however. In order to make sure that funds are indeed available in case of a medical need, spending from the account on nonmedical expenditures must be prohibited. However, in order for cost-sharing to have an incentive effect, the money spent on cost-sharing requirements must have an opportunity cost. If spending from an HSA is confined to medical expenditures, the cost of today's spending is leaving possible future medical needs unattended. This opportunity cost may turn out too vague to create much of an incentive to spend the money cost-consciously. A transfer into an HSA thus presents a dilemma for policy-makers between committing the recipient and offering incentives for cost-conscious spending.[19]

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 pro-

[18]For overviews on their effectiveness, see Buchmueller (2008) and Bundorf (2016).

[19]The current practice in the United States is to penalize nonmedical spending from the account. It is unclear whether this compromise between imperfect commitment and incentivization leads to a more efficient utilization given the difficulty to discern the actual opportunity cost of medical spending (Buchmueller, 2008; Bundorf, 2016).

ducing 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 ones. 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 actual prices in a negative incentive scheme.

ACKNOWLEDGEMENTS

I thank Friedrich Breyer, the participants of the CEAR/MRIC Workshop on Behavioral Insurance 2017, the NOeG Annual Meeting 2018, the European Health Economics Workshop 2019, and the seminar audience at Karlsruhe Institute of Technology and University of Dortmund (TU) for valuable comments.

ORCID

Markus Fels https://orcid.org/0000-0002-0404-2094

REFERENCES

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

Besanko, D., Dranove, D., & Garthwaite, C. (2016). Insurance and the high prices of pharmaceuticals. NBER Working Paper No. 22353.

Boone, J. (2018). Basic versus supplementary health insurance: Access to care and the role of cost effectiveness. *Journal of Health Economics*, *60*, 53–74.

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

Brot-Goldberg, Z. C., Chandra, A., Handel, B. R., & Kolstad, J. T. (2017). What do deductibles do? The impact of cost-sharing on health care prices, quantities, and spending dynamics. *The Quarterly Journal of Economics*, *132*, 1261–1318.

Buchmueller, T. C. (2008). Consumer-oriented health care reform strategies: A review of the evidence on managed competition and consumer-directed health insurance. *The Milbank Quarterly*, *87*(4), 820–841.

Bundorf, M. K. (2016). Consumer-directed health plans: A review of the evidence. *The Journal of Risk and Insurance*, *83*(1), 9–41.

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

Cutler, D. M., & Zeckhauser, R. J. (2000). Chapter 11 - The Anatomy of Health Insurance. In A. Culyer and J. Newhouse (Eds.), *Handbook of Health Economics* (Vol. *IA*, pp. 563–643). Amsterdam: Elsevier.

Dionne, G., Fombaron, N., & Doherty, N. (2013). Adverse selection in insurance contracting. In G. Dionne (Ed.), *Handbook of insurance* (pp. 231–280). New York: Springer-Verlag.

Fels, M. (2019). Risk attitudes with state-dependent indivisibilities in consumption. Ruhr Economic Papers No. 805.

Fels, M. (2020). Mental accounting, access motives, and overinsurance. *Scandinavian Journal of Economics*, *122*(2), 675–701.

Fels, M., Kairies-Schwarz, N., & Waibel, C. (2020). Cost-sharing or rebate? What drives patients' treatment decisions? https://papers.ssrn.com/abstract=3545781, Working Paper.

Fendrick, A. M., & Chernew, M. E. (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. M., Smith, D. G., Chernew, M. E., & Shah, S. N. (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.

Hayen, A. P., Klein, T. J., & Salm, M. (2018). Does the framing of patient cost-sharing incentives matter? the effects of deductibles vs. no-claim refunds. IZA Discussion Paper No. 11508.

Jaspersen, J. G., & Richter, A. (2015). The wealth effects of premium subsidies on moral hazard in insurance markets. *European Economic Review*, *77*, 139–153.

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

Kunreuther, H. C., Pauly, M. V., & McMorrow, S. (2013). *Insurance & behavioral economics: Improving decisions in the most misunderstood industry*. Cambridge: Cambridge University Press.

McGuire, T. G. (2011). Chapter Five - Demand for Health Insuranceemand for Health Insurance. In M.V. Pauly, T.G. McGuire and P.P. Barros (Eds.), *Handbookof Health Economics* (Vol. *2*, pp. 317–396). Amsterdam: Elsevier.

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

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

Nyman, J. A. (2003). *The Theory of Demand for Health Insurance*. Stanford: Stanford University Press.

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

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

Remmerswaal, M., Boone, J., Bijlsma, M., & Douven, R. (2019). Cost-sharing design matters: A comparison of the rebate and deductible in health care. *Journal of Public Economics*, *170*, 83–97.

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

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

Schoen, C., Osborn, R., Squires, D., Doty, M. M., Pierson, R., & Applebaum, S. (2010). How health insurance design affects access to care and costs, by income, in eleven countries. *Health Affairs*, *29*(12), 2323–2334.

Schokkaert, E., Steel, J., & Van de Voorde, C. (2017). Out-of-pocket payments and subjective unmet need of healthcare. *Applied Health Economics and Health Policy*, *15*(5), 1–11.

Siu, A. L., Sonnenberg, F. A., Manning, W. G., Goldberg, G. A., Bloomfield, E. S., Newhouse, J. P., & Brook, R. H. (1986). Inappropriate use of hospitals in a randomized trial of health insurance plans. *The New England Journal of Medicine*, *315*(20), 1259–1266.

Tamblyn, R., Laprise, R., Hanley, J. A., Abrahamowicz, M., Scott, S., Mayo, N., ..., & Mallet, L. (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., & Manning, W. G. (2000). Chapter 8 - Moral Hazard and Consumer Incentives in Health Care. In A.J. Culyer and J.P. Newhouse (Eds.), *Handbook of Health Economics* (Vol. *1A*, pp. 409–459). Amsterdam: Elsevier.

How to cite this article: Fels M. Incentivizing efficient utilization without reducing access: The case against cost-sharing in insurance. *Health Economics*. 2020;29:827–840. https://doi.org/10.1002/hec.4023