Cost incentives for doctors:
A double-edged sword*

Christoph Schottmüller†

February 28, 2013

Abstract

If doctors take the costs of treatment into account when prescribing medication, their objectives differ from their patients' objectives because the patients are insured. This misalignment of interests hampers communication between patient and doctor. Giving cost incentives to doctors increases welfare if (i) the doctor's examination technology is sufficiently good or (ii) (marginal) costs of treatment are high enough. If the planner can costlessly choose the extent to which doctors take costs into account, he will opt for less than 100%. Optimal health care systems should implement different degrees of cost incentives depending on type of disease and/or doctor.

JEL: D82, D83, I10

Keywords: cheap talk, patient-doctor communication, health insurance, health market design

1 Introduction

It is well known that insurance creates moral hazard: In the health sector, insured people would like to have more expensive treatments than socially optimal. On the other hand, treatments are normally prescribed by doctors. If doctors took the costs of treatment into account in their treatment decision, the moral hazard problem should disappear. The tradition in the medical profession, however, is to view oneself as advocate of one's patients. Consequently, the patient's well-being is put first and costs are only secondary. What is more, doctors are often explicitly hostile towards cost incentives in doctor remuneration. The German chamber of doctors, for instance, writes in its principles of health policy[1]

> [. . .] the role of the doctor as advocate for his patient must not be restricted [. . .] The state must not establish financial schemes (e.g. bonus-malus system) which could suggest to the patient that materialistic, self-serving aspects are also of importance for medical decisions.

It is important to understand whether the doctors' concerns are mainly self-interested, e.g. worries about reputation and pay, or whether financial incentives for doctors could have a negative impact on social welfare. Put differently, can patient advocacy be interpreted as an efficient institutional response to the particular structure of the health care market? Answering this question will also give some insight into the optimal design of health care markets. In particular, in which parts of the health care system should cost incentives for doctors be employed and where are cost incentives less likely to succeed?

This paper focuses on the communication between patient and doctor. The patient's input, e.g. describing his symptoms and their intensity, is vital to

---

[1]Translation by the author. Original title and source: "Gesundheitspolitische Leitsätze der Ärzteschaft–Ulmer Papier" Beschluss des deutschen Ärztetags 2008, Anlage 1, p. 6, `http://www.bundesaerztekammer.de/downloads/UlmerPapierDAET111.pdf`

reach the right diagnosis.[2] The main mechanism I explore in this paper is the following: Patients are (fully) insured. If doctors take costs into account in their treatment decision, their objectives and the objectives of their patients are no longer aligned.[3] Such a misalignment undermines the patient's trust in his doctor which in turn affects communication negatively.[4] More technically, in a setting where the patient has private information, e.g. about his symptoms and their intensity, he has the possibility to exaggerate his symptoms (or their intensity) in order to get a more expensive treatment. Of course, the doctor will anticipate such strategic exaggerating. This anticipation gives the patient further incentives to exaggerate and so on.

The appropriate model to analyze such a "rat race" is the cheap talk framework. This paper will therefore extend the canonical cheap talk model to the imperfect information setting typical for the health sector. Although a complete breakdown of communication can be prevented, communication will be worse in equilibrium because of the misalignment of interests, i.e. less information is transmitted from patient to doctor. It is shown that this communication effect can make a system without cost incentives preferable from a social welfare point of view. If the patient's collaboration is hardly needed, a system with cost incentives is preferable. For example, a doctor can easily establish that a patient has a broken leg by having an X-ray. The symptoms reported

---

[2]The importance of communication is also stressed in the aforementioned document of the German chamber of doctors where it is stated that "health can neither be commanded nor produced since health depends crucially on the patient's collaboration." Also there is a whole string of the medical literature dealing with doctor-patient communication, see Stewart (1995) for a survey.

[3]Negative effects from cost incentives on the doctor-patient relationship are also established in the medical literature, see for example Rodwin (1995), Kao et al. (1998) or Gallagher and Levinson (2004).

[4]There is no doubt that patients understand this nexus: According to Gallagher et al. (2001) 73% of their respondents dislike the idea of a cost control bonus for their doctor and 91% favor disclosure to the patient if such a bonus was in place. Furthermore, 95% of those who dislike the bonus stated that the bonus would lower their trust in their physician.

3

by the patient are less important in this case. If, on the other hand, an illness might have a psychological background, the patient's collaboration is essential and a system without cost incentives might be preferable.

From a technical point of view, the paper contributes to the cheap talk literature following the seminal paper by Crawford and Sobel (1982). Their model is extended in Chen (2009) and de Barreda (2010) to a setup where the decision maker receives a noisy signal. My paper generalizes further by substituting the perfect information on the sender/expert/patient side by a noisy signal.[5]

This paper complements existing literature on the design of health care systems. Early contributions as Arrow (1963) and Pauly (1968) already point out the moral hazard caused by health insurance: Insured patients might over-consume treatment from a social welfare perspective because they are insured. Ma and McGuire (1997) introduce the physician as an additional player and analyze contractual difficulties in the health market. In particular, health outcome and doctor's effort are non-contractible and even the quantity of care consumed can be subject to misreporting. Ma and McGuire (1997) analyze how these contractual constraints influence optimal contracts between insurance and patient as well as between insurance and physician. My paper focuses on a different kind of constraint, i.e. a constraint in information transmission arising in the communication between doctor and patient. It will be shown that the necessity of information transmission between patient and doctor might constrain the power of the incentive scheme offered to the doctor.

Obviously related is the literature on physician compensation and managed care. In his survey of the managed care literature, Glied (2000) mentions two problems of "supply-side cost sharing," i.e. cost incentives for physicians: (i) underprovision of necessary services and (ii) strong incentives to avoid costly

---

[5]Ishida and Shimizu (2010) also considers a setting where both sides have noisy signals. They consider the case where the state of the world is binary and the signal space is discrete. My paper uses a continuum of health states and signals.

cases. In this context, my paper adds a third problem: Hampered information transmission between doctor and patient. Furthermore, my paper provides one possible explanation for the ambiguous cost effect of managed care mentioned in Glied (2000).

Also related is the literature on physician agency with asymmetric information, see McGuire (2000) for a survey. However, this literature focuses mainly on the observability and contractibility of quality and effort choices while my paper analyzes communication between doctor and patient. An exception to this focus is the literature on supply induced demand, see Pitchik and Schotter (1987); Calcott (1999); De Jaegher and Jegers (2001). These papers model a doctor sending cheap talk messages concerning recommended treatments to the patient. A conflict of interest emerges as the doctor maximizes his income and not patient utility. To the best of my knowledge, my paper is the first one to model communication from the patient to the doctor.

The medical literature contains statements like "payment arrangements could significantly undermine patients' beliefs that their physicians are acting as their agents" (Mechanic and Schlesinger, 1996) and emphasizes that there should be no conflict of interest between patient and doctor (Emanuel and Dubler, 1995).[6] Kao et al. (1998) find that patients trust their physician less if the physician is capitated than when he is payed on a fee for service basis.[7] Physicians are also less satisfied with their relationships with capi-

---

[6] See McGuire (2000) for more references on this point. The focus of these papers differs slightly from my paper as they concentrate on doctor's own income maximization as a reason for mistrust and diverging interests. I will abstract from this and focus direclty on the discrepancy between welfare and patient utility caused by health insurance.

[7] It should be noted that doctors' and patients' incentives are also not aligned under a fee for service arrangement as a doctor has incentives to overtreat the patient, see the discussion in section 2. However, patients appear to be less worried about overtreatment in practice. The reasons might be that many insurance plans actively try to prevent costly overtreatment, e.g. by utilization reviews, and also that patients do not bear the financial risk of overtreatment because of insurance. Therefore, objectives of doctor and patient are normally viewed to be closer in a fee-for-service contract.

tated patients compared to their average patient (Kerr et al., 1997). My paper contributes by formalizing why trust, interpreted as shared objectives, is vital for the patient-physician relationship. Such a formalization is interesting for two reasons: First, it allows for both costs (less trust) and benefits (less overtreatment) of cost incentives. Second, one can obtain results concerning the optimal design of health care systems, i.e. where in the health system are aligned interests especially important and where could cost incentives improve welfare.

The next section introduces the model and is followed by a simple numerical example. This example illustrates the main points. Section 4 analyzes a general model and answers the question: When do cost incentives work? Two extensions are analyzed in section 5: In the first one, the planner can costlessly choose any arbitrary degree to which the doctor should take costs into account. It is shown that the optimal degree is less than 100%. The second extension analyzes how copayments can help to alleviate the communication problem. The final section concludes by discussing the results and pointing out predictions as well as possible applications in different areas. Proofs are relegated to the appendix.

2 Formal setting

Patient and doctor have a common prior F over the set of all possible health states of the patient. The set of health states is denoted by Θ. The patient receives a private signal $\sigma^p \in \Sigma^p$ about his health state. In practice this signal can be interpreted as the symptoms a patient can report to his doctor or as the intensity of his symptoms. The doctor receives also a private signal $\sigma^d \in \Sigma^d$ about the patient's health. This signal can be interpreted as the result of the doctor's examination, e.g. his interpretation of an X-ray photograph or listening to the patient's heartbeat. Given the health state, there is a

distribution $G(\sigma^p, \sigma^d|\theta)$ of signals which is common knowledge. Put differently, $G(\sigma^p, \sigma^d|\theta)$ gives the probabilities that a patient (doctor) receives signal σ^p (σ^d) given a health state θ.

The timing is the following: First, the patient's health state is determined by nature. This health state is unknown to doctor and patient. Second, doctor and patient receive their signals $\sigma = (\sigma^p, \sigma^d)$ which correspond to the true health state through G. Third, the patient can send a message, e.g. communicating his signal, to the doctor. Fourth, the doctor determines a treatment τ from a set of available treatments. The costs of the treatment $c(\tau)$ are paid for by the patient's insurance.

Utility of the patient depends only on his true health state $\theta \in \Theta$ and the treatment τ. In particular, a patient's well being does in the end not depend on the signals. For the doctor, I look at two scenarios: Either the doctor has "no cost incentives" which means that he makes his treatment decision to maximize the patient's utility or he is "cost sensitive" (or "has cost incentives") with which I mean that he maximizes social welfare. Social welfare is the patient's utility minus costs. The perspective of the paper is therefore eventually the perspective of a (benevolent) designer of the health system, e.g. a government or an insurance plan, who has to determine which kind of incentives he gives to the doctor.[8]

I want to discuss briefly under which contract forms a doctor might maximize patient utility or welfare. This issue will then be neglected in the remainder as it is not the main focus of the paper. While it is hardly disputed that doctors care about their patient's health, it is likewise undisputed that doctors react to financial incentives, see Armour et al. (2001); Brook (2010) or McGuire (2000). Financial incentives can be of two kinds. First, capitation

---

[8]It will become clear that–very much in line with this design perspective–the cost function can include more than just the costs of medication, e.g. social costs due to absence at work or risk of infection for others. If one wants to allow for partial insurance, the cost function $c(\tau)$ would then be the part of costs that are not borne by the insured himself.

payments, cost saving bonuses and similar schemes give an incentive to save costs, i.e. to reduce τ in the model. Second, fee-for-service arrangements give an incentive to overtreat the patient as additional treatment leads to more fees. This would be interpreted as an incentive to increase τ in the model. A situation with–roughly–no financial incentives is the case of a salaried doctor. All mentioned forms of enumeration (and combinations of those) are used in practice, see Brook (2010). Assume that a doctor cares about his patient's utility with weight $\alpha > 0$ and about his income with weight 1. A doctor will then maximize welfare if his contract consists of a fixed payment minus $\alpha c(\tau)$. A doctor maximizes patient utility if he receives a fixed payment only.

3 A simple example

This section deals with a small numerical example which illustrates that cost incentives can lead to lower welfare. Take $\Theta = \{A, B, C\}$ and $\Sigma^p = \Sigma^d = \{0, 1\}$. In words, there are three diseases called A, B and C. Doctor and patient will each receive one of two possible signals which are denoted by 0 and 1. For example, the patient's signal could be whether he feels "no/little pain" or "strong pain" while the doctor's signal could be whether the patient's heartbeat is unusual or not. The prior F is given by disease A and B occurring with probability 2/5 each and disease C with probability 1/5. The distribution G is given in the following table:

prior	2/5	2/5	1/5
σ	A	B	C
(0,0)	0	0	1
(0,1)	0	4/5	0
(1,0)	1/5	1/5	0
(1,1)	4/5	0	0

The interpretation is that, given health state A, signal $(\sigma^p, \sigma^d) = (1, 1)$ occurs with probability 4/5 and signal $(\sigma^p, \sigma^d) = (1, 0)$ occurs with probability

1/5. Assume that there are three available treatments which are denoted by a, b and c. The patient's utility and the costs of each treatment are given in the following table:

	A	B	C	costs
a	8	9.7	9.2	5
b	4	9	9.6	3
c	0	5	10	1

To illustrate: A patient with disease A receiving treatment a has a utility of 8. Treatment a costs 5. Therefore, welfare would be $8 - 5 = 3$ in this situation.

One interpretation is that "disease" C is being healthy and treatment c is the option "no additional treatment" (the costs of 1 would be the costs of the initial doctor visit). Treatment a is a very effective and expensive treatment while b is a less effective and cheaper treatment. Overtreatment reduces utility slightly. A quick calculation shows that treatment a is welfare maximizing in health state A where welfare is defined by patient utility minus costs. The same is true for b in health state B and c in C.[9]

3.1 No cost incentives

If the doctor has no cost incentives, the incentives of doctor and patient are aligned. The patient will therefore communicate his true signal σ^p in equilibrium.[10] The doctor can then base his decision on both signals and maximizes

---

[9]The example is a discretization of the model in section 4: The health states θ corresponding to A, B and C are $10, 5, 0$. The treatments a, b, c correspond to values of τ of $8, 4, 0$. The utility function leading to the values above is

$$u(\theta - \tau) = \begin{cases} -(\theta - \tau) + 10 & \text{if } \theta - \tau \geq 0 \\ (\theta - \tau)/10 + 10 & \text{if } \theta - \tau < 0. \end{cases}$$

[10]In principle, there is also a pooling equilibrium in which the doctor takes only his own signal into account and the patient sends the same message regardless of his signal. However,

gross consumer surplus. Hence, the doctor knows the disease whenever the signals are $(0,0)$, $(0,1)$ or $(1,1)$. If the signal is $(1,0)$, the doctor assigns equal probabilities to disease A and B. This leads to the following optimal decisions: $(0,0) \rightarrow c$, $(0,1) \rightarrow a$, $(1,0) \rightarrow a$ and $(1,1) \rightarrow a$

Expected welfare is therefore[11]

$$W^{nci} = \frac{1}{5}(10-1) + \frac{8}{25}(9.7-5) + \frac{2}{25}(8-5) + \frac{2}{25}(9.7-5) + \frac{8}{25}(8-5) = \frac{122}{25}.$$

3.2 Cost sensitive doctor

If the doctor is cost sensitive, his preferred decisions (if he knew both signals) would be: $(0,0) \rightarrow c$, $(0,1) \rightarrow b$, $(1,0) \rightarrow a$ and $(1,1) \rightarrow a$. Hence, there is a conflict between the patient and the doctor whenever the signal is $(0,1)$: The doctor prefers treatment b while the patient prefers a. Next, I write down the optimal decision of the doctor if he only knows his own signal σ^d. If $\sigma^d = 1$, he assigns equal probability to disease A and B. Therefore, the optimal treatment is a. If $\sigma^d = 0$, he assigns probability $2/9$ to disease A, $2/9$ to disease B and $5/9$ to disease C. It is straightforward to calculate that in this case the optimal treatment is c.

In principle, there could be two kinds of equilibrium: First, a separating equilibrium in which the patient truthfully reports his signal to the doctor, i.e. the two signals are separated. Second, a pooling equilibrium in which the patient sends the same message regardless of his signal.[12]

Suppose there is a separating equilibrium, i.e. the patient communicates his signal σ^p truthfully to the doctor in equilibrium. The doctor will then

---

this equilibrium is Pareto dominated and does not seem very realistic.

[11]Just to illustrate: The first term is the probability of being in state C and receiving the signal $(0,0)$, i.e. $1/5 * 1$, multiplied with the utility of the resulting treatment c in state C, i.e. 10, minus the costs of this treatment, i.e. 1.

[12] It can be shown that there is no mixed strategy equilibrium with an outcome different from the pooling outcome.

implement the welfare maximizing treatment knowing both signals. If $\sigma^p = 0$, the patient expects–given his signal–to get a utility of $u^{truth} = 8/13*9+5/13*10 = 122/13$.[13] If however the patient lied and communicated $\sigma^p = 1$, the doctor would implement treatment a and the patient's expected utility would be $u^{lie} = 8/13*9.7+5/13*9.2 = 1236/130$. Hence, lying pays off for the agent and there cannot be a separating equilibrium.

Consequently, there is a pooling equilibrium in which the doctor uses only his own signal. Welfare is then

$$W^c = \frac{1}{5}(10-1) + \frac{8}{25}(9.7-5) + \frac{2}{25}(0-1) + \frac{2}{25}(5-1) + \frac{8}{25}(8-5) = \frac{1206}{250}.$$

Since $W^c < W^{nci}$, cost incentives reduce welfare in this example. The driving force behind this result are the conflicting objectives of patient and doctor which result in a break down of communication. Nevertheless, costs are lower if the doctor is cost sensitive since the signal $(1,0)$ leads to the low cost treatment c while a is prescribed without cost incentives.

3.3 Variation I: Restricting the choice set

Interestingly, there is an easy fix in this example: Suppose, the health authority does not clear treatment b. Hence, treatment b is not available. But then there is no conflict between doctor and patient as even a cost sensitive doctor will now prescribe a if the signal $(0,1)$ occurs. Unfortunately, this means that cost incentives simply do not matter/work: Every signal leads to the same treatment with and without cost incentives.[14] Furthermore, this trick will not always work: Amend the example above with a disease D which can be identified with certainty (so there would be a signal $(2,2)$ which occurs if and

---

[13]Given $\sigma^p = 0$, the patient assigns probability 8/13 to health state B with signal $\sigma = (0,1)$ which leads to treatment b. With the counter probability 5/13, he expects state C with signal $\sigma = (0,0)$ and treatment c.

[14]This point is more general and also holds in the model analyzed in section 4 if one adds the possibility to restrict the treatment set.

only if the health state is D). If in this state D treatment b is by far superior to all other treatments, a health authority banning treatment b would reduce welfare.

3.4   Variation II: Increasing costs

The negative information effect of cost incentives can be so strong that costs can be higher under cost incentives. To see this, change the example above by changing the ex ante probability of disease C from $1/5$ to $p^c < 1/5$ and assign the ex ante probability $(1 - p^c)/2$ to sickness A and B. Note that this does not change decisions without cost incentives as it is always perfectly known whether one is in state C or not.

If, however, p^c is small enough and the doctor knows only his own signal, he will prescribe treatment a instead of treatment c (or b) when he receives signal $\sigma^d = 0$. This inevitably leads to higher costs than without cost incentives: Now a is always prescribed while c was prescribed without cost incentives for signal $\sigma = (0, 1)$. Note that a lower p^c will make the incentive constraint of a separating equilibrium even tougher, i.e. reducing p^c does not lead to a separating equilibrium. It turns out that in the example a is the optimal treatment for $\sigma^d = 0$ if $p^c < 0.029$, i.e. if $p^c < 0.029$ costs with cost incentives are higher than without.

This result is slightly reminiscent of the empirical results concerning the cost effects of managed care. One feature of many managed care plans are cost incentives for doctors, e.g. cost control boni or capitation payment. As Glied (2000) reports in his survey, results on the cost effect of managed care are however inconclusive: Some studies report higher costs, some report lower costs or no cost difference between managed care and traditional care plans.

4   Model and results

This section uses a more general model to analyze the setting and effect described before. There are two reasons why this is desirable: First, one has to verify that the effects described above are not due to the discrete nature of the example. Second, this will allow to determine under which circumstances cost incentives are welfare maximizing and therefore have implications for the optimal design of a health care system.

The patient's message in the example above is "cheap talk": The message itself does not have direct payoff implications. Only the treatment decision is relevant for the patient's utility and welfare. The canonical model for cheap talk games is Crawford and Sobel (1982). To fit the health sector, the information structure of Crawford and Sobel (1982) has to be amended as described below.

I assume that health state θ is a real number from some bounded interval and also σ^p, σ^d and τ are assumed to be real numbers.[15]   Without loss of generality take $\Theta = [0,1]$. Higher signals are assumed to imply higher expected states. To make this formal define by $H(\theta | \sigma^p, \sigma^d)$ the cumulative distribution function which gives the probability that the state is below θ given signals σ^p and σ^d. This distribution is derived from $F(\theta)$ and $G(\sigma^p, \sigma^d | \theta)$ using Bayes' rule. The assumption is that $H(\theta | \sigma^p, \sigma^d)$ first order stochastically dominates $H(\theta | \sigma^{p\prime}, \sigma^{d\prime})$ whenever $\sigma^d \geq \sigma^{d\prime}$ and $\sigma^p \geq \sigma^{p\prime}$. In words, a higher signal implies that higher health states are more likely to occur.[16]

Patient utility $u(\theta - \tau)$ is a function of "difference" between health state and treatment. It is assumed that the patient is fully insured, i.e. costs of treatment do not enter his utility function. Assume that $u(\theta - \tau)$ is two

---

[15]Restricting τ to some interval, e.g. $\mathbb{R}_+$ is possible as explained in footnote 23. Drawing the signals from some closed subset of $\mathbb{R}$ simplifies matters, see assumption 1.

[16] If G has a density g which is differentiable in σ, the stochastic dominance assumption can be written as $\frac{\int_0^\theta g_{\sigma^i}(\sigma | \theta)\, dF(\theta)}{\int_0^\theta g(\sigma | \theta)\, dF(\theta)} \leq \frac{\int_0^1 g_{\sigma^i}(\sigma | \theta)\, dF(\theta)}{\int_0^1 g(\sigma | \theta)\, dF(\theta)}$ for each $i \in \{p, d\}$, $\theta \in \Theta$, $\sigma \in \Sigma^p \times \Sigma^d$.

times continuously differentiable, strictly concave and attains its maximum at 0. Put differently, patient utility is maximized if $\tau = \theta$ and is lower the further away treatment τ is from this ideal treatment. A treatment above (below) θ corresponds to overtreatment (undertreatment) from the patient's point of view. It is *not* assumed that $u(\cdot)$ is symmetric and therefore over- and undertreatment might affect utility in different ways. The cost function $c(\tau)$ is strictly increasing and marginal costs are bounded away from 0, i.e. $c'(\tau) \geq \delta \quad \forall \tau$ for some $\delta > 0$. This last assumption implies that the patient's utility is never aligned with the social objective or, put differently, the patient always prefers a more expensive treatment than socially optimal because he is insured. If there was no such conflict, cost incentives would simply not matter for the outcome. Consequently, introducing cost incentives could not even help to reduce costs.

The solution concept is Perfect Bayesian Nash Equilibrium. After observing his signal σ^p a patient updates his beliefs about his health state θ and about the doctor's signal. Given σ^p, a strategy for the patient is a probability distribution over Σ^p denoted by $q(m|\sigma^p)$.[17] This distribution gives the probability of reporting $m \in \Sigma^p$ when the true signal is σ^p. For illustration purposes, think of a partition equilibrium in which patients with signals in, say, $[0.3, 0.4]$ are bunched, i.e. send the same message. In this case $q(m|\sigma^p)$ could be a uniform distribution over $[0.3, 0.4]$ for all $\sigma^p \in [0.3, 0.4]$. Given his signal σ^d and the message he receives from the patient, the doctor updates his beliefs about the health state of the patient θ and chooses his preferred treatment. For simplicity, I assume that $u(\theta - \tau) - c(\tau)$ is strictly concave in τ which implies that there is a unique socially efficient treatment τ^w. This assumption is, for example, satisfied if $c(\tau)$ is linear or convex. Hence, the doctor will always have a unique preferred treatment which I denote by $\tau^d(m, \sigma^d)$. The strategies $(q(m|\sigma^p), \tau^d(m, \sigma^d))$ form an equilibrium if:

---

[17]For notational convenience $q(m|\sigma^p)$ is a probability density function but mass points can be easily accommodated.

1. For each σ^p, $q(m|\sigma^p)$ is a distribution, i.e. $\int_0^1 q(m|\sigma^p)\,dm = 1$, and if $q(m^*|\sigma^p) > 0$ then $m^* \in argmax_m \int_0^1 \int_{\Sigma^d} u(\theta - \tau^d(m, \sigma^d))\,dP(\theta, \sigma^d|\sigma^p)$ where $P(\theta, \sigma^d|\sigma^p)$ is the distribution of (θ, σ^d) derived from $G(\sigma^p, \sigma^d|\theta)$ and $F(\theta)$ conditional on observing σ^p and using Bayes' rule.[18]

2. For each m and σ^d, treatment maximizes the doctor's objective. For the cost sensitive doctor, this means that $\tau^d(m, \sigma^d) = argmax_\tau \int_0^1 [u(\theta - \tau) - c(\tau)]\,dH(\theta|m, \sigma^d)$ where with a slight abuse of notation $H(\theta|m, \sigma^d)$ is the distribution of the health state conditional on observing σ^d and m using by Bayes' rule (given $G(\sigma^p, \sigma^d|\theta)$, $F(\theta)$ and $q(m|\sigma^p)$). Without cost incentives $\tau^d(m, \sigma^d) = argmax_\tau \int_0^1 u(\theta - \tau)\,dH(\theta|m, \sigma^d)$.

In words, the first condition says that the patient reports with positive probability only signals maximizing his utility given the strategy of the doctor. The second condition establishes that the doctor uses an optimal strategy given the patient's equilibrium behavior.

I define a *monotone partition equilibrium* as an equilibrium characterized by a partition $\{s_0, s_1, \ldots, s_n\}$ of Σ^d such that (i) $q(m|\sigma^p) = q(m|\sigma^{p\prime})$ if and only if σ^p and $\sigma^{p\prime}$ belong to the same element of the partition and (ii) the support of $q(m|\sigma^p)$ and $q(m|\sigma^{p\prime})$ is non overlapping if σ^p and $\sigma^{p\prime}$ belong to different elements of the partition.

The focus of the paper is on monotone equilibria as non-monotone partition equilibria, as analyzed in Chen (2009), appear unnatural in patient doctor communication. Put differently, it is easy to imagine that patients who observe symptoms for two or three days send the same message. However, it is hard to imagine that these patients send the same message as patients observing symptoms for three weeks while patients with one or two weeks send a different message. However, the results of this section still hold if non-monotone partition equilibria exist.

---

[18]Note that the patient takes expectations not only over the health state but also over the doctor's signal because σ^d will influence the doctor's treatment decision.

The following technical assumption proves to be useful for the analysis. Note that the boundedness part is automatically satisfied if H_{σ^p} is continuous and the signal σ is drawn from a closed set, i.e. if Σ^p and Σ^d are closed intervals.

Assumption 1. *$H(\theta|\sigma^p, \sigma^d)$ is differentiable in σ^p and $|H_{\sigma^p}(\theta|\sigma^p, \sigma^d)|$ is bounded from above by some $M > 0$. At all states where $H(\theta|\sigma^p, \sigma^d)$ has a density $h(\theta|\sigma^p, \sigma^d)$, this density is also differentiable in σ^p and h_{σ^p} is bounded.*[19]

Put differently, beliefs about the true health state do not change too sharply if the patient's signal changes marginally. Note that slightly irregular distribution, e.g. with mass points at a "healthy state" $\theta = 0$, can be allowed. Assumption 1 simplifies the analysis by ensuring that the doctor's treatment decision is differentiable in the patient's signal. In fact, it implies that there is an upper bound on how strongly the doctor's treatment decision reacts to a marginal change in σ^p (in a hypothetical situation in which the doctor knows the patient's signal). Loosely speaking, this means that a patient who exaggerates his signal a little bit will–as a consequence–only get a slightly higher treatment. See the proof of theorem 1 for details.

The game is then similar to the information transmission model of Crawford and Sobel (1982) with three additional twists: First, the doctor (receiver in the language of Crawford and Sobel) receives a signal while he is completely ignorant in Crawford and Sobel (1982). Second, the patient (sender) does not know the state of the world. Instead, he has a noisy signal. Third, the divergence of interests between doctor (receiver) and patient (sender) is not fixed but depends on the treatment (decision). The following theorem extends results from Crawford and Sobel (1982) to this setting.

---

[19] This assumptions is satisfied if G has a density g which is differentiable in σ^p and g_{σ^p} is bounded.

Theorem 1. *With cost incentives, there exists no separating equilibrium. Monotone partition equilibria exist. All but the first element of the partition have a minimum length κ which is bounded away from zero. If Σ^p is bounded, the number of elements in the partition is bounded from above.*

⁵ **Proof.** see appendix A.1

The intuition is the following: In equilibrium, a patient cannot tell his true signal to the doctor. If he did, the doctor would prescribe a treatment that is "too cheap" from the patient's point of view (as the patient does not care about costs). Hence, the patient would have an incentive to overstate ¹⁰ his signal. In practice, this would mean to claim additional symptoms or to overstate the intensity of existing symptoms. What happens in equilibrium is that the patient's signal range is partitioned and the patient reports in which element of the partition his signal lies. The doctor does not know the precise signal of the patient but gets a rough idea which he takes into consideration ¹⁵ when choosing the treatment. Because of the partitioning, a patient can no longer overstate his signal "a little bit". If the patient deviated by reporting a higher element of the partition, he would get a substantially higher treatment. In equilibrium he will not deviate because he expects this treatment to be too high. One could interpret this in the following two ways: First, a patient does ²⁰ not want to report symptoms that are too much different from the real ones as this could mislead the doctor, i.e. result in treating the wrong illness. Second, extreme overstatement of symptoms could result in too strong medication with severe side effects. Hence, the patient does not want to overstate his existing symptoms too much.

²⁵ It is also clear that the partition cannot be arbitrarily fine: If the elements are too small, then overstating one's signal "a little bit" is again possible. This explains the minimum length statement in the theorem. The minimum element length immediately implies that the number of elements is bounded if the interval from which patient signals are drawn is bounded.

The mechanism through which cost incentives can harm welfare is the same as in the example of section 3: If the objectives of doctor and patient are different, the patient has an incentive to use his information strategically to get the more expensive treatment he wants. In equilibrium, the doctor will have less information (partitioning of signal range) compared to the situation without cost incentives. Consequently, he is more prone to make inappropriate treatment decisions. In short, there are two effects when introducing cost incentives: First, costs are taken into account which, ceteris paribus, decreases costs and increases welfare. Put differently, the doctor stops prescribing excessively expensive treatments. Second, communication and therefore the information of the doctor is worse. Hence, treatment decisions are less accurate which reduces welfare. Whether the cost or the information effect dominates is ex ante unclear. The following propositions show that in two extreme cases the cost effect dominates and therefore cost incentives lead to higher welfare than no cost incentives.

Proposition 1. *Welfare is higher with cost incentives if the doctor's signal is sufficiently informative. That is, given $G(\sigma^p, \sigma^d | \theta)$, for $\varepsilon > 0$ small enough cost incentives lead to higher welfare than no cost incentives if the doctor's signal is drawn from $\varepsilon G(\sigma^p, \sigma^d | \theta) + (1 - \varepsilon) 1_\theta$ where 1_θ is a distribution putting all probability mass on θ. Cost incentives lead also to higher welfare if the patient's signal is sufficiently uninformative, i.e. for $\varepsilon > 0$ small enough if the patient's signal is drawn from $\varepsilon G(\sigma^p, \sigma^d | \theta) + (1 - \varepsilon) U_\theta$ where U_θ is the uniform distribution over $[0, 1]$.*

Proof. see appendix A.1

This result is intuitive: If the doctor is able to determine the patient's health state almost on his own, i.e. without knowing the patient's signal, then the patient's signal is useless. Therefore, the information effect of introducing cost incentives is small while the cost effect is still there.

One interpretation of proposition 1 is that cost incentives become eventually

18

more attractive with medical progress. This holds at least true if medical progress implies better diagnosis possibilities for doctors. Consequently, one might then expect to see more cost incentive elements in health care systems over time.

A second interpretation is that some specialists optimally should have cost incentives while others should not. A radiologist or a trauma surgeon will normally base his decisions on his own examination and less on the patient's report. This might be less true for an internist or a general practitioner.

A related third interpretation is that an optimal health care system should incorporate selective cost incentives. More precisely, cost incentives should be applied for the treatment of diseases where the doctor's information is relatively more important than the patient's information.

Proposition 2. *Cost incentives lead to higher welfare than no cost incentives if social and private objectives differ sufficiently. That is, for any given information structure and cost function $c(\tau)$ there exists an $\alpha > 0$ such that cost incentives lead to higher welfare than no cost incentives under the cost function $\alpha c(\tau)$.*

Proof. see appendix A.1

The intuition is that the cost effect will become dominant if (marginal) costs are high enough. Consequently, the information loss due to cost incentives is negligible compared to the cost effect.

In line with previous interpretations cost incentives are especially useful for specialists dealing with high cost treatments on a regular basis. Also diseases involving high cost treatment on a regular basis are especially well suited for cost incentives.

To conclude this section, I want to point out that the tradeoff between information and cost effect is also present in simpler models. Put differently, the model structure with noisy signals for patient and doctor reflects the reality in the health care sector but is not necessary to generate the result that no

cost incentives can be optimal for welfare. To verify this claim, I show that no cost incentives are welfare optimal in the archetypical cheap talk example introduced by Crawford and Sobel (1982).

Example. *Health states are uniformly distributed on $[0, 1]$. The patient has perfect knowledge of the health state while the doctor's signal is completely uninformative. Assume that the patient's utility function is a quadratic loss function, i.e. $u(\theta, \tau) = -(\theta - \tau)^2$, and that the cost function is linear in treatment, i.e. $c(\tau) = \alpha\tau$. Given the information that σ^p (which is now the true health state) is in the interval (s_1, s_2), the optimal treatment decision for a doctor with cost incentives is $\tau = \frac{s_1 + s_2 - \alpha}{2}$. With $\alpha = 1/10$ the model is equivalent to the example in Crawford and Sobel (1982). It is shown there that the finest possible equilibrium partition is $(0, 2/15, 7/15, 1)$, i.e. a patient will report whether his signal is in $[0, 2/15)$ or in $[2/15, 7/15)$ or in $[7/15, 1]$. Straightforward calculations show that expected consumer utility in this partition equilibrium is -0.01058 while expected costs are 0.045. Hence, expected welfare is $-0.01058 - 0.045 = -0.05558$.*

Without cost incentives the patient will truthfully reveal his signal and therefore communicate the true health state to the doctor. Consequently, $\tau = \theta$ and consumer welfare is 0. Expected costs are $\frac{1}{10}0.5 = 0.05$ which results in expected welfare of -0.05. Therefore, no cost incentives lead to higher welfare than cost incentives.

5   Extensions

This section considers two extensions. In these extensions, I will concentrate on the finest monotone partition equilibrium. To ensure uniqueness of the finest equilibrium partition, a standard monotonicity condition is assumed which is explained in detail in appendix A.2.

5.1 Degree of cost incentives

Say, the planner could set the precise *extent* to which the doctor takes costs into account: The planner sets $\beta \in [0, 1]$ and the doctor maximizes the expected value of $u(\theta - \tau) - \beta c(\tau)$ with his treatment decision. The following theorem says that neither $\beta = 1$ nor $\beta = 0$ will be optimal in this setting. Hence, a welfare maximizing planner does not want a welfare maximizing doctor.

Theorem 2. *Assume that the doctor's signal is not perfect. Then the optimal degree of cost incentives is interior, i.e. the optimal β is neither 0 nor 1.*

Proof. see appendix A.2

The idea is the following: Say $\beta = 1$ induces the partition $\{s_0, s_1, \ldots, s_N\}$. Now suppose β is decreased marginally (starting from $\beta = 1$) and assume for now that the partition remained the same: Then the doctor would prescribe higher treatments in response. As he was a welfare maximizer before this change, this will only have a second order effect on welfare (if the partition did not change). However, there will be a first order information effect: Because the doctor prescribes higher treatments, the interests of patient and doctor differ less. Hence, the equilibrium partition will be finer. Roughly speaking, the patient trusts the doctor more when β is decreased and is therefore willing to transmit more information. For $\beta = 0$, the argument works in the opposite direction: There is no first order information effect but a first order cost saving effect from increasing β.

Note that there is a second interpretation of the first/second order argument above: Take some fixed $\beta > 0$. The argument above says that a doctor maximizing $u - \beta c$ would like to commit to a lower β, say $\tilde{\beta} < \beta$. Such a commitment would allow the doctor to achieve a higher value of $u - \beta c$. However, any such commitment is difficult because the doctor's signal is not observable. Put differently, after receiving the patient's message the doctor has an incentive to implement the treatment τ maximizing $u - \beta c$. He can then always claim that τ would maximize $u - \tilde{\beta} c$ but that he received a lower

signal than he actually did, i.e. the patient will not even realize the deviation. Nevertheless, one could interpret doctors' emphasis of patient advocacy and even the Hippocratic oath as attempts to commit to a low $\tilde{\beta}$.

5.2   Demand side cost sharing

Copayments and deductibles are commonly used instruments in health insurance. The standard argument for the use of these instruments is moral hazard, i.e. patients might overdemand health care without copayments. This argument ignores the fact that most care has to be prescribed by a doctor whose incentives might differ from the patient's preferences. Put differently, this argument assumes that the doctor follows the patient's wishes in his prescription behavior.

The model of this paper suggests that the moral hazard argument might be valid in an indirect way: Copayments imply that the patient takes costs partially into account. Objectives of patient and cost sensitive doctor can therefore be better aligned with copayments. This will improve communication thereby increasing welfare. The mechanism is similar to the moral hazard argument as the copayments reduce the preferred treatment of the patient. However, the welfare improvement stems not directly from this demand reduction but from the effect it has on communication. If the doctor does not have cost incentives and acts in the patient's interest, the standard moral hazard argument applies directly.

Obviously, copayments have the downside of exposing the patient to financial risk. As risk aversion is the reason for the existence of insurance in the first place, copayments cannot be too high. The optimal level of copayments has to balance this negative risk effect with the positive effect on communication.

To demonstrate these ideas in the model, a financial dimension has to be added. Assume that the patient has utility $v(w - p - \gamma c(\tau)) + u(\theta - \tau)$ where v is an increasing and concave function, w is income, p is the insurance premium

22

and γ is the copayment rate.[20] We are interested in the structure of the welfare maximizing insurance contract under the constraint that the insurance breaks even in expectation, i.e. $p = (1 - \gamma)\,\mathbb{E}[c(\tau)]$. This insurance contract is in itself interesting for normative reasons. Furthermore, this contract would be offered in the equilibrium of a perfectly competitive insurance market. The following proposition assumes the same monotonicity condition as theorem 2 and confirms the intuition above.

Proposition 3. *Copayments are strictly positive in the welfare maximizing insurance contract, i.e. $\gamma > 0$.*

Proof. see appendix A.2

It should be noted that theorem 2 still holds also in this setup. Optimally, health markets should therefore include copayments and give doctors less than full cost incenitves.

5.3 Two-sided communication

Chen (2009) introduces two sided communication–the doctor can send a cheap talk message to the patient before the patient communicates his message–in a setting in which the patient knows the state with certainty and the doctor has a binary signal. In this setting, there can be equilibria with meaningful communication from the doctor to the patient. Intuitively, a doctor with a high signal wants to communicate this as the patient will be less afraid of undertreatment. In certain circumstances, Chen (2009) shows that also a doctor with a low signal might want to report the low signal truthfully (roughly speaking this is true if reporting a high signal would change the information parition too much).

In the setting of this paper, where the patient has a noisy signal, there is an additional effect of two-sided communication. A doctor communicating a

---

[20]For simplicity, a fixed copayment rate that does not depend on treatment is used here. See appendix A.2 for extending the result below to more flexible copayment schemes.

high signal does not only communicate that he will prescribe a high treatment but also that high states are likely. Put differently, the message of the doctor contains information not only about his prescription behavior but also about the state. The two kinds of transmitted information affect the patient's will-
ingness to communicate in opposite ways: If the patient believes that higher states are more likely, he has more incentives to exaggerate his signal. If he believes that the doctor will prescribe a high treatment, he has less incentives to exaggerate.

While a complete characterization of the two-sided communication case is
beyond the scope of this paper, it seems intuitive that the previous results still hold in a two-sided communication framework. This is obvious when the doctor cannot communicate truthfully in equilibrium. If partial communication from doctor to patient is possible, the message of the doctor will simply update the beliefs of the patient. Then a one-sided communication situation similar to the
one analyzed in this paper emerges as a subgame. This observation implies, for example, that theorem 1 and propositions 1 and 2 hold in the two-sided communication game because they hold in any subgame following any message by the doctor. The proofs of theorem 2 and proposition 3 will go through as long as (i) the monotonicity condition holds and (ii) a marginal change in the
parameter (β and γ respectively) does not change the communication from the doctor to the patient in a discontinuous fashion.

6   Discussion and conclusion

Introducing cost incentives for doctors turns out to be a double-edged sword: On the one hand, taking costs into consideration should avoid the prescription
of too expensive treatments. On the other hand, misalignment of patient's and doctor's incentives will hamper communication between the two: The patient has an incentive to exaggerate and in equilibrium this leads to signal bunching. Consequently, the doctor has worse information and is less likely

24

to assess the patient's health state correctly. Knowing about the uncertainty he might even choose more expensive treatments to be on the safe side. In a numerical example, this can lead to higher costs than under no cost incentives (see section 3).

If costs are very high or if the doctor is able to assess the health state very accurately given only his signal, cost incentives are the welfare maximizing policy. This shows that an optimal health care system will use different degrees of cost incentives in different circumstances. In practice, cost incentives could differ across diseases and across specialists. However, it is shown that full cost incentives, i.e. doctors take all costs into account, are not optimal. Copayments can help to mitigate the communication problem as they bring the objectives of cost sensitive doctor and patient closer together. Although copayments have the obvious disadvantage of exposing the risk averse patient to risk, they are strictly positive in the welfare maximizing insurance contract.

The model can also be interpreted as a formalization of the idea that trust is important in the patient-doctor relationship. This idea is commonplace in the medical literature and was informally discussed in Arrow (1963). Surprisingly, the health economics has largely ignored this topic since then. As in definitions of trust, see Bhattacharya et al. (1998), the model describes a situation where one person (patient) relies on the future action of another person (doctor). Trust can then be defined as a feeling of confidence of the former (patient) that the latter (doctor) will act in his (patient) interest. Intuitively, a patient should be confident that the doctor acts in his interest if both share the same objectives.[21] Therefore, cost incentives reduce trust in the patient-doctor relationship. In this interpretation, the model gives one explanation why trust is important in medical care: Less trust corresponds to worse communication, less information transmission and a worse diagnosis. Put differently, the

---

[21]In the classification of Gilson (2003), this is trust in the *strategic* perspective. The alternative *altruistic* perspective views trust as an institution enabling cooperation in situation where a party could profitably deviate from cooperation.

information effect identified earlier can be interpreted as a *trust effect.*

In some sense, the model is a best case scenario for the benevolent designer: He can freely set the doctor's incentives without incurring any costs. In practice, setting up an incentive scheme for doctors might actually be costly. Doctors might also not respond immediately because of previously formed habits. It is therefore even more remarkable that the designer might not want to give cost incentives to the doctor in the model of this paper (and sets less than full cost incentives when he can choose the precise extent of cost incentives).

The model gives several predictions. Quality of diagnosis should decrease after an introduction of cost incentives for doctors: For example, patients with a given diagnosis-treatment pair will be treated less successfully (e.g. take longer to recover) because some receive the wrong treatment due to a wrong diagnosis. This effect should be more pronounced for specialists and diseases where patient input is vital for the diagnosis. If trust reflects the willingness to communicate, one should expect patient's trust in their doctor to be lower when their doctor has cost incentives. This last result is indeed confirmed by the empirical health literature, see for example Kao et al. (1998).

More abstract, a welfare maximizing sponsor (say a benevolent government) might prefer a decision maker (doctor) who shares his preferences not with the sponsor but with the patient. In a broader context an agent might benefit from surrendering his interests when information provision by another party is important. This could have applications in other contexts like mediation: A mediator with decision power who shares the interests of another party might be preferable to making the decision oneself.

In general, shared objectives prove to be vital for information provision. Patient advocacy can therefore be seen as an institutional response to the importance of information provision by patients. Consequently, one might expect similar institutions to emerge whenever information provision by affected parties is vital. In this context, the relationship between a lawyer and his client could serve as an additional example.

A Appendix

A.1 Proofs

Proof of theorem 1: The proof proceeds in a number of steps. The first three steps establish that there cannot be a separating equilibrium, i.e. there is no equilibrium in which a patient always reports his true signal. Consequently, patients with some signals are bunched together. Patients in one "bunch" (one element of a partition of the signal range) send the same report to the doctor. Steps four and five establish that each element of a partition must have a minimum length, i.e. the partition cannot be arbitrarily fine.

The first step is to show that there exists a $b > 0$ such that $argmax_\tau \int_0^1 [u(\theta - \tau) - c(\tau)] \, dH(\theta|m, \sigma^d) + b \leq argmax_\tau \int_0^1 u(\theta - \tau) \, dH(\theta|m, \sigma^d)$ for a given equilibrium strategy $q(m|\sigma^p)$; i.e. the patient would opt for an at least b higher treatment than a cost sensitive doctor if he chose (and had the same information). This follows from the first order conditions corresponding to the two $argmax$ expressions

$$\int_0^1 -u'(\theta - \tau) \, dH(\theta|m, \sigma^d) = \begin{cases} c'(\tau) \\ 0 \end{cases} . \tag{1}$$

The left hand side of (1) is continuous in τ and also strictly decreasing in τ. Since $c'(\tau) \geq \delta > 0$ and $u'(\cdot)$ is continuous, the claim follows. This argument is for a given (m, σ^d) but the infimum of all these b over (m, σ^d) will also be strictly positive. To establish this, it is sufficient to show that the derivative of the left hand side of (1) with respect to τ is bounded:[22] Since $u'(\theta - x) > 0$ for $x \geq 1$ and any $\theta \in [0, 1]$, the optimal treatment is bounded from above by 1. Furthermore, the optimal treatment is bounded from below by $\underline{\tau}$ solving $u'(-\underline{\tau}) = c'(\underline{\tau})$, i.e. the optimal treatment if the doctor knew that $\theta = 0$.

---

[22]Just to illustrate why boundedness is sufficient: Say the derivative of the left hand side of (1) is between 0 and $-B$. Since this left hand side is differentiable, the two τ solving (1) with the right hand side equal to zero and equal to $c'(\tau)$ have to differ by at least δ/B.

Therefore $-1 \leq \theta - \tau \leq 1 - \underline{\tau}$. By the continuity of $u''(\cdot)$ and the compactness of $[-1, 1 - \underline{\tau}]$, $u''(\cdot)$ is bounded on this interval. Consequently, the derivative of the left hand side of (1) is a weighted (by the distribution $H(\cdot)$) average of a bounded function and therefore bounded. Denote by $B > 0$ such a bound on the derivative of the left hand side of (1). Then we can choose $b = \delta/B$.[23]

Second, the patient prefers a slightly higher treatment than a cost sensitive doctor prescribes in a hypothetical separating equilibrium. From the first step and the strict concavity of $u(\cdot)$, it follows that any treatment in $(\tau^d, \tau^d + b)$ yields a higher expected utility for the patient than τ^d.

Third, in a hypothetical separating equilibrium the patient attains a higher utility by misrepresenting slightly upwards as the doctor will increase his decision uniformly continuously in σ^p. The implicit function theorem gives for a hypothetical separating equilibrium

$$\frac{d\tau^d}{d\sigma^p} = \frac{\frac{\partial \int_0^1 -u'(\theta-\tau)\, dH(\theta|\sigma^p,\sigma^d)}{\partial \sigma^p}}{-\int_0^1 [u''(\theta-\tau) - c''(\tau)]\, dH(\theta|\sigma^p,\sigma^d)}. \tag{2}$$

The denominator is obviously positive as it is (-1) times the second order condition of the doctor's maximization problem. The numerator is positive as well because of stochastic dominance: As $-u'(\theta - \tau)$ is a strictly increasing function of θ, we have $\int_0^1 -u'(\theta - \tau)\, dH_1(\theta) > \int_0^1 -u'(\theta - \tau)\, dH_2(\theta)$ whenever $H_1(\theta)$ first order stochastically dominates $H_2(\theta)$. Since $H(\theta|\sigma^{p\prime},\sigma^d)$ first order stochastically dominates $H(\theta|\sigma^p,\sigma^d)$ whenever $\sigma^{p\prime} > \sigma^p$, the numerator has to be positive. The uniform continuity follows from the boundedness of 2: The numerator is bounded by assumption 1 and the fact that $u'(\theta - \tau)$ is bounded on the relevant range. The strict concavity of the doctor's program implies that the denominator is strictly bounded away from zero.[24] By uniform continuity,

---

[23]If the treatment is restricted to be larger than, say, 0, the argument still holds true as long as $H(0|0,0) < 1$. A patient will then always desire a treatment that is strictly bounded away from 0. Therefore, interests of patient and doctor are not aligned even if the constraint $\tau \geq 0$ is binding.

[24]To be precise, this follows as the treatment range is bounded by $\underline{\tau}$ and 1. On this closed

misrepresentation can be chosen small enough to prevent an "overreaction" by the doctor.

Consequently, there cannot be a separating equilibrium. The same argument shows that also locally, i.e. on some subinterval of the patient's signal range, there cannot be a perfect separation of types, i.e. patient signals have to be bunched in equilibrium.

Fourth, in a partition equilibrium communicating a higher partition element will result in a higher treatment decision. This follows from the fact that higher signals σ^p indicate higher health states θ and the doctor's optimal treatment decision is increasing in θ. Formally speaking, $H(\theta|(s_1, s_2), \sigma^d)$ first order stochastically dominates $H(\theta|(s_1', s_2'), \sigma^d)$ whenever $s_1' < s_2' \leq s_1 < s_2$.

Fifth, in a partition equilibrium there exists a minimum length $\kappa > 0$ of each (but the first) partition element. It was shown earlier that the optimal treatment decision of a doctor is uniform continuous in σ^p (in a hypothetical separating equilibrium). Therefore, there exists a $\kappa > 0$ such that optimal treatment decisions differ by less than b for all σ^p and $\sigma^{p\prime}$ with $|\sigma^p - \sigma^{p\prime}| < \kappa$ (in a hypothetical separating equilibrium). Now suppose by way of contradiction that there was a partition element (s_0, s_1) with $s_1 - s_0 < \kappa$. By the definition of κ and b, a patient with signal $\sigma^p = s_0$ will (in expectation) strictly prefer the cost sensitive doctor's separating treatment decision for type $\sigma^p = s_1$ to the separating treatment decision for type $\sigma^p = s_0$. By concavity of $u(\cdot)$, he will also prefer a cost sensitive doctor's separating treatment decision for all types $\sigma^p \in (s_0, s_1)$ to his own. By continuity, the same holds for patients with a signal $s_0 - \varepsilon$ for some $\varepsilon > 0$ small enough. Clearly, a cost sensitive doctor receiving the message (s_0, s_1) will assign a treatment between the optimal separating treatment for $\sigma^p = s_0$ and for $\sigma^p = s_1$. Therefore, a patient with signal $s_0 - \varepsilon$ will prefer the message (s_0, s_1) to any message $m \subset [0, s_0]$.

Step five and boundedness of the patient's signal range imply that the

_____

and bounded treatment range the maximum of the second derivative exists and constitutes the bound away from 0.

number of partitions in any partition equilibrium is bounded.

A one-element-partition equilibrium ("babbling equilibrium") in which all σ^p are pooled exists always. This proves existence of partition equilibria. $\square$

Proof of proposition 1: Denote the doctor's beliefs over states θ (derived by Bayes' rule) given a signal drawn from $\varepsilon G(\sigma^p, \sigma^d | \theta) + (1 - \varepsilon)1_\theta$ by $k(\theta, \varepsilon | \sigma^d)$. Note that these beliefs are continuous in ε. For $\varepsilon = 0$, the doctor has full information and therefore the welfare maximum is attained with cost incentives. As $c'(\tau) > 0$, decisions under no cost incentives differ from decisions with cost incentives. Consequently, welfare with cost incentives is strictly higher than without cost incentives if $\varepsilon = 0$. As beliefs (and therefore treatment decisions and welfare) are continuous in ε, it follows that even in the babbling equilibrium welfare with cost incentives is higher than welfare without cost incentives for $\varepsilon > 0$ small enough. With cost incentives welfare in any partition equilibrium will be at least as high as in the babbling equilibrium. This implies the first part of the proposition.

For the second part, note that $H(\theta | \sigma^p, \sigma^d)$ does not depend on σ^p if $\varepsilon = 0$. Consequently, no information is lost when switching to cost incentives. Taking costs into account makes cost incentives strictly superior as $c'(\tau) > 0$. By continuity of $H(\theta | \sigma^p, \sigma^d)$ in ε, the same conclusion holds for $\varepsilon > 0$ small enough (for the babbling equilibrium and therefore even more so for other partition equilibria). $\square$

Proof of proposition 2: Since $c'(\tau) \geq \delta > 0$, there exists an α such that

$$-u'(1) - \alpha c'(0) \leq 0.$$

This implies that the welfare maximizing treatment decision τ is non-positive for any signal/message under the cost function $\alpha c(\tau)$. Without cost incentives $\tau \geq 0$ and $\tau(\sigma^p, \sigma^d) > 0$ with strictly positive probability as

$$\int_0^1 -u'(\theta) \, dH(\theta | \sigma^p, \sigma^d) > 0$$

whenever $H(0 | \sigma^p, \sigma^d) < 1$. Consequently, welfare is lower without cost incentives compared to the simple policy $\tau = 0$ (regardless of the signal) under cost

function $\alpha c(\tau)$. A cost sensitive doctor will improve on this simple policy by using the information he has, i.e. σ^d. Consequently, cost incentives lead to higher welfare than no cost incentives under the cost function $\alpha c(\tau)$. $\qquad\square$

A.2   Proofs Extensions

This section gives the proofs for theorem 2 and proposition 3. I assume for simplicity that patient's signals are drawn from the interval $[0,1]$ and that G and F have strictly positive densities. To make comparisons between different levels of cost incentives meaningful, this section assumes that the *finest equilibrium partition* is the market outcome. In an equilibrium partition $\{0, s_1, \ldots, s_{n-1}, 1\}$, each s_i with $i = 1, \ldots, n-1$ has to be indifferent between the elements (s_{i-1}, s_i) and (s_i, s_{i+1}).[25] Using this indifference condition and the value s_1, one can calculate s_2. Given s_2 one calculates s_3 using the indifference condition and so on. Following Crawford and Sobel (1982), I will call the result of this calculation procedure starting with s_1 a *forward solution*. Equivalently, one could start from s_{n-1} and calculate s_{n-2} using the indifference condition. Then one continues with s_{n-3} etc.. The result of this will be called a *backward solution*.

In this section, I use the monotonicity condition that was introduced in Crawford and Sobel (1982) and is often used in the cheap talk literature, see Chen (2009) for a recent example in a related framework.[26]

(M) For a given cost function and β, if s and $\tilde{s}$ are two forward solutions with $s_0 = \tilde{s}_0 = 0$ and $s_1 > \tilde{s}_1$, then $s_i > \tilde{s}_i$ for all $i \geq 2$.

(M') For a given cost function and β, if s and $\tilde{s}$ are two backward solutions with $s_n = \tilde{s}_{\tilde{n}} = 1$ and $s_{n-1} > \tilde{s}_{\tilde{n}-1}$, then $s_{n-i} > \tilde{s}_{\tilde{n}-i}$ for all $i \geq 2$.

This regularity condition ensures that for a given n, there is at most one equilibrium partition with n elements. This means that the *finest equilibrium*

---

[25]To shorten notation I write "each s_i" instead of "a patient with a signal $\sigma^p = s_i$ for some $i = 1 \ldots n-1$".

[26]The following arguments bear some similarity with theorems 3-5 in Crawford and Sobel (1982) which are shown under a similar monotonicity condition.

partition is uniquely defined.

A.2.1 Interior cost incentives are optimal (Proof of theorem 2)

This subsection formally shows that a social planner optimally chooses an interior level of cost incentives.

So, a setting where the planner can set β and the doctor maximizes $u - \beta c$ is analyzed. It will be shown that the optimal β is below one.

To analyze whether $\beta = 1$ is optimal, it is first necessary to determine how the partition changes when β is decreased.

Lemma 1. *If β is decreased and the number of elements in the equilibrium partition does not change, then s_i increases for all $i = 1 \ldots, n - 1$.*

Proof. A lower β implies higher treatment τ for a given partition element and doctor signal. Hence, a patient with a given signal in a given partition element will expect a higher treatment if β is lower. Take $\beta^l < \beta^h$ with corresponding equilibrium partitions $\{0, s_1^j, \ldots, s_{n-1}^j, 1\}$ where $j = h, l$.

The first result is the following: If $s_i^l \leq s_i^h$ for some i, then the same holds for all smaller i. Take i as the highest $i < n$ where $s_i^l \leq s_i^h$. Now, let us work backwards: s_i^l is indifferent between (s_{i-1}^l, s_i^l) and (s_i^l, s_{i+1}^l). The proof moves from the indifference condition of s_i^l to the indifference condition of s_i^h in three steps: (i) changing s_{i+1}^l to s_{i+1}^h, (ii) changing from s_i^l to s_i^h and (iii) changing from β^l to β^h. In all steps, the lower bound of the lower partition element–s_{i-1}^l at the beginning–has to be increased to keep indifference. This shows that $s_{i-1}^l < s_{i-1}^h$. The statement above follows then by induction.

(i) Since $s_{i+1}^l > s_{i+1}^h$, s_i^l will prefer (s_i^l, s_{i+1}^h) over (s_{i-1}^l, s_i^l).[27] Therefore, there is an $\hat{s}_i > s_{i-1}$ such that s_i is indifferent between $(\hat{s}_i, s_i^l)$ and (s_i^l, s_{i+1}^h).

---

[27]Intuitively, s_i is indifferent between (s_{i-1}^l, s_i^l) and (s_i^l, s_{i+1}^l) because he is undertreated in the former and overtreated in the latter (from his own point of view). Reducing the upper bound of the higher interval reduces the "overtreatment" and makes it more attractive for s_i.

(ii) By (M'), there is a $\check{s}_i \geq \hat{s}_i$ such that s_i^h is indifferent between $(\check{s}_i, s_i^h)$ and (s_i^h, s_{i+1}^h) (under β^l!). (iii) Last, change β^l to β^h. As this decreases treatment, s_i^h will prefer (s_i^h, s_{i+1}^h) over $(\check{s}_i, s_i^h)$. Hence, $s_{i-1}^h > \check{s}_i$.

The second result is the following: If $s_1^l \leq s_1^h$, then $s_i^l < s_i^h$ for all $i = 2, \ldots, n-1$. There are two steps to prove this: We start at the indifference condition of s_1^l. Then (i) s_1^l is changed to s_1^h and (ii) β^l is changed to β^h. In both steps, the upper bound of the upper partition element–which is s_2^l at the start–has to be increased to keep indifference. Hence, $s_2^h > s_2^l$ and the result follows by induction.

(i) Under β^l, s_1^l is indifferent between $(0, s_1^l)$ and (s_1^l, s_2^l). As $s_1^h \geq s_1^l$, (M) implies that there is a $\hat{s}_2 \geq s_2^l$ such that s_1^h is indifferent between $(0, s_1^h)$ and $(s_1^h, \hat{s}_2)$ under β^l. (ii) Under β^h treatment is lower and higher partition elements are therefore ceteris paribus more attractive. Hence, s_2^h such that s_1^h is indifferent between $(0, s_1^h)$ and (s_1^h, s_2^h) under β^h has to satisfy $s_2^h > \hat{s}_2$. Therefore, $s_2^h > s_2^l$.

The previous results imply that whenever $s_i^h > s_i^l$ for some i, then $s_{n-1}^h > s_{n-1}^l$. To show that $s_i^h \leq s_i^l$, it is therefore sufficient to show $s_{n-1}^h \leq s_{n-1}^l$. This last part is a proof by contradiction. So, suppose $s_{n-1}^h > s_{n-1}^l$. By the first result, $s_1^h > s_1^l$. Then by (M), there is a $\bar{s}_1 > s_1^l$ such that a forward solution starting at $\bar{s}_1$ yields $\bar{s}_{n-1} = s_{n-1}^h$ (under β^l). Note that by (M), $\bar{s}_i > s_i^l$. This implies that $\bar{s}_{n-1}$ strictly prefers $(\bar{s}_{n-1}, 1)$ to $(\bar{s}_{n-2}, \bar{s}_{n-1})$ (since a $\bar{s}_n$ satisfying the indifference condition "would have to be above 1").

By the second result, $\bar{s}_1 > s_1^h$ as otherwise $\bar{s}_{n-1} < s_{n-1}^h$ which contradicts the definition of $\bar{s}$. Since $\bar{s}_{n-1} = s_{n-1}^h$, $\bar{s}_{n-2} \geq s_{n-2}^h$: Otherwise, according to result 1 above $\bar{s}_1 > s_1^h$ could not hold. Consequently, s_{n-1}^h prefers $(\bar{s}_{n-2}, s_{n-1}^h)$ over $(s_{n-1}^h, 1)$ under β^h. This is the same as saying $\bar{s}_{n-1}$ prefers $(\bar{s}_{n-2}, \bar{s}_{n-1})$ over $(\bar{s}_{n-1}, 1)$ under β^h. Under β^l lower partition elements become more attractive and therefore the same holds true under β^l. But this contradicts the conclusion of the last paragraph. Hence, $s_{n-1}^h > s_{n-1}^l$ cannot hold and $s_i^h \leq s_i^l$ is true. $\qquad \square$

Write expected welfare as

$$W = \int_{\Sigma^d} \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} \int_0^1 \left[u(\theta - \tau(s_{i-1}, s_i, \sigma^d)) - c(\tau(s_{i-1}, s_i, \sigma^d)) \right]$$

$$dH(\theta|\sigma^p, \sigma^d) \, prob(\sigma^p|\sigma^d) d\sigma^p \, dProb(\sigma^d)$$

where $prob(\sigma^p|\sigma^d)$ is the ex ante density of σ^p given σ^d derived from F and G using Bayes' rule. In the same way $Prob(\sigma^d)$ is the ex ante unconditional distribution of σ^d. Note that all s_i and τ depend on β. Now take the derivative of welfare with respect to β at $\beta = 1$. Since the doctor maximizes expected welfare on each partition element, an envelope argument yields that $\frac{d\tau}{d\beta}$ can be neglected. Hence,

$$\frac{dW}{d\beta}\bigg|_{\beta=1} = \int_{\Sigma^d} \sum_{i=1}^{n} \left(\int_0^1 \left[u(\theta - \tau(s_{i-1}, s_i, \sigma^d)) - c(\tau(s_{i-1}, s_i, \sigma^d)) \right] dH(\theta|s_i, \sigma^d) \right.$$

$$\left. - \int_0^1 \left[u(\theta - \tau(s_i, s_{i+1}, \sigma^d)) - c(\tau(s_i, s_{i+1}, \sigma^d)) \right] dH(\theta|s_i, \sigma^d) \right) prob(s_i|\sigma^d) \frac{ds_i}{d\beta} \, dProb(\sigma^d).$$

According to the lemma above $ds_i/d\beta < 0$. Because of that and since $c(\tau(s_{i-1}, s_i)) < c(\tau(s_i, s_{i+1}))$, leaving out the cost terms will increase the right hand side of the previous equation. Hence,

$$\frac{dW}{d\beta}\bigg|_{\beta=1} < \int_{\Sigma^d} \sum_{i=1}^{n} \left(\int_0^1 \left[u(\theta - \tau(s_{i-1}, s_i, \sigma^d)) - u(\theta - \tau(s_i, s_{i+1}, \sigma^d)) \right] dH(\theta|s_i, \sigma^d) \right)$$

$$prob(s_i|\sigma^d) \frac{ds_i}{d\beta} \, dProb(\sigma^d)$$

$$= \sum_{i=1}^{n} \left\{ \int_{\Sigma^d} \int_0^1 \left[u(\theta - \tau(s_{i-1}, s_i, \sigma^d)) \right. \right.$$

$$\left. \left. - u(\theta - \tau(s_i, s_{i+1}, \sigma^d)) \right] prob(\sigma^d, \theta|s_i) \, d\theta \, d\sigma^d \right\} prob(s_i) \frac{ds_i}{d\beta}$$

$$= 0.$$

The last equality follows from the indifference condition which holds for every s_i. The indifference condition states that the term in curly brackets is 0 for each i. Consequently, $dW/d\beta$ is negative at $\beta = 1$ and it is welfare improving to lower β below one.

Similarly, $dW/d\beta$ is positive at $\beta = 0$ and therefore the optimal β is interior. To see this, note that the finest partition for $\beta = 0$ is fully separating.[28] Hence, $\tau(s_{i-1}, s_i, \sigma^d)$ approaches $\tau(s_i, s_{i+1}, \sigma^d)$ as $\beta \to 0$ (eventually being $\tau(s_i, s_i, \sigma^d)$ in the limit). Therefore, the $ds_i/d\beta$ are multiplied by zero terms in the $dW/d\beta$ expression and drop out. Consequently,

$$
\left. \frac{dW}{d\beta} \right|_{\beta=0} = \int_{\Sigma^d} \int_{\Sigma^p} \int_0^1 \left(-u'(\theta - \tau(\sigma^p, \sigma^p, \sigma^d)) - c'(\tau(\sigma^p, \sigma^p, \sigma^d)) \right)
$$
$$
H(\theta | \sigma^p, \sigma^d) \frac{d\tau(\sigma^p, \sigma^p, \sigma^d)}{d\beta} \, prob(\sigma^p | \sigma^d) \, d\sigma^p \, dProb(\sigma^d).
$$

As the doctor maximizes the expected patient utility (with $\beta = 0$), the term $\int_0^1 u'(\theta - \tau(\sigma^p, \sigma^p, \sigma^d)) dH(\theta | \sigma^p, \sigma^d)$ is zero. Since $-c' < 0$ and $d\tau/d\beta < 0$, $dW/d\beta > 0$ at $\beta = 0$ follows.

A.2.2 Copayments are optimal (Proof of proposition 3)

First, I want to analyze the case where the doctor is cost sensitive.[29] With copayments equal to zero the model collapses to the model of theorem 1. The proof of proposition 3 is by contradiction. Suppose there is a treatment τ prescribed in equilibrium with positive probability and $\gamma = 0$. It will be shown that marginally increasing γ increases welfare in this case. Expected welfare in this setting is

$$
W = \int_{\Sigma^d} \sum_{i=1}^n \int_{s_{i-1}}^{s_i} \int_0^1 v(w - p - \gamma c(\tau(s_{i-1}, s_1, \sigma^d))) + u(\theta - \tau(s_{i-1}, s_1, \sigma^d))
$$
$$
dH(\theta | \sigma^p, \sigma^d) \, prob(\sigma^p | \sigma^d) d\sigma^p \, dProb(\sigma^d)
$$

where by assumption

$$
p = (1-\gamma)\mathbb{E}c(\tau) = (1-\gamma) \int_{\Sigma^d} \sum_{i=1}^n \int_{s_{i-1}}^{s_i} c(\tau(s_{i-1}, s_1, \sigma^d)) \, prob(\sigma^p | \sigma^d) d\sigma^p \, dProb(\sigma^d)
$$

---

[28]It is straightforward toextent the argument of Agastya et al. (2012) that the finest distribution converges to full separation to the framework of this paper.

[29]The proof still goes through if the doctor is partially cost sensitive, i.e. has a β in $(0, 1]$.

in equilibrium. Insurance profits can be neglected when denoting welfare as they are zero by assumption.

Increasing γ has a direct effect on welfare and an indirect effect by changing the equilibrium information partition. As a first step, it is shown that the direct effect of marginally increasing γ is zero when $\gamma = 0$. The idea is that the patient does not face financial risk when there is no coinsurance. Consequently, there is no first order welfare effect from transferring financial risk to the patient through copayments.

$$
\begin{aligned}
\left.\frac{\partial W}{\partial \gamma}\right|_{\gamma=0} &= \int_{\Sigma^d} \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} \int_0^1 \left[(1 - v'(w-p)) \left(\frac{\partial p}{\partial \gamma} + c(\tau(s_{i-1}, s_1, \sigma^d)) \right) \right] \\
&\quad dH(\theta|\sigma^p, \sigma^d) \; prob(\sigma^p|\sigma^d) d\sigma^p \; dProb(\sigma^d) \\
&= (1 - v'(w-p)) \int_{\Sigma^d} \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} \int_0^1 \left(-\mathbb{E}c(\tau) + c(\tau(s_{i-1}, s_1, \sigma^d)) \right) \\
&\quad dH(\theta|\sigma^p, \sigma^d) \; prob(\sigma^p|\sigma^d) d\sigma^p \; dProb(\sigma^d) \\
&= 0
\end{aligned}
$$

Hence, it remains to show that the indirect effect of copayments, i.e. the effect through changing the information partition, increases welfare at $\gamma = 0$. The following lemma shows that information is improved when γ increases.

Lemma 2. *If γ is increased and the number of partition elements remains the same, then s_i increases for all $i = 1, \ldots, n-1$.*

Proof. First, note that a higher γ reduces the most preferred treatment τ. Hence, higher γ reduces the wedge between cost sensitive doctor's and patient's interest. In this sense, a high γ is similar to a low β in the previous subsection. The proof of this lemma is similar to the proof of lemma 1 and therefore only sketched.

Take $\gamma^l < \gamma^h$ with corresponding equilibrium partitions $\{0, s_1^j, \ldots, s_{n-1}^j, 1\}$ where $j = h, l$.

The first result is the following: If $s_i^h \leq s_i^l$ for some i, then the same holds for all smaller i. Let i the highest $i < n$ where $s_i^h \leq s_i^l$. s_i^h is indifferent between

(s_{i-1}^h, s_i^h) and (s_i^h, s_{i+1}^h). The proof moves from the indifference condition of s_i^h to the indifference condition of s_i^l in three steps: (i) changing s_{i+1}^h to s_{i+1}^l, (ii) changing from s_i^h to s_i^l and (iii) changing from γ^h to γ^l. In all three steps, the lower bound of the lower partition element–s_{i-1}^h at the beginning–has to be increased to keep the indifference condition. This shows that $s_{i-1}^h < s_{i-1}^l$. The statement above follows then by induction.

The second result is the following: If $s_1^h \leq s_1^l$, then $s_i^h < s_i^l$ for all $i = 2, \ldots, n-1$. This is proven in two steps: Take the indifference condition for s_1^h. Then (i) s_1^h is changed to s_1^l and (ii) γ^h is changed to γ^l. In both steps, the upper bound of the upper partition element–which is s_2^h at the start–has to be increased to keep the indifference condition. Hence, $s_2^l > s_2^h$ and the result follows by induction.

The previous results imply that whenever $s_i^l > s_i^h$ for some i, then $s_{n-1}^l > s_{n-1}^h$. To show that $s_i^l \leq s_i^h$, it is consequently sufficient to show $s_{n-1}^l \leq s_{n-1}^h$. This part of the proof is by contradiction. Suppose $s_{n-1}^l > s_{n-1}^h$. By the first result, $s_1^l > s_1^h$. Then by (M), there is a $\bar{s}_1 > s_1^h$ such that a forward solution starting at $\bar{s}_1$ yields $\bar{s}_{n-1} = s_{n-1}^l$ (under γ^h). Note that by (M), $\bar{s}_i > s_i^h$. Therefore, $\bar{s}_{n-1}$ strictly prefers $(\bar{s}_{n-1}, 1)$ to $(\bar{s}_{n-2}, \bar{s}_{n-1})$ (since a $\bar{s}_n$ satisfying the indifference condition "would have to be greater than 1").

By the second result, $\bar{s}_1 > s_1^l$ as otherwise $\bar{s}_{n-1} < s_{n-1}^h$ contradicting the definition of $\bar{s}$. As $\bar{s}_{n-1} = s_{n-1}^l$, $\bar{s}_{n-2} \geq s_{n-2}^l$: Otherwise, $\bar{s}_1 > s_1^l$ could not hold because of the first result above. Hence, s_{n-1}^l prefers $(\bar{s}_{n-2}, s_{n-1}^l)$ over $(s_{n-1}^l, 1)$ under γ^l. This is the same as saying $\bar{s}_{n-1}$ prefers $(\bar{s}_{n-2}, \bar{s}_{n-1})$ over $(\bar{s}_{n-1}, 1)$ under γ^l. Under γ^h lower partition elements become more attractive and therefore the previous statement is also true under γ^h. But this contradicts the conclusion of the last paragraph. Hence, $s_{n-1}^l > s_{n-1}^h$ cannot hold and therefore $s_i^l \leq s_i^h$. $\qquad\qquad\square$

Using the result above that the direct effect of a change in γ on welfare is

zero (at $\gamma = 0$), the total effect on welfare can be written as

$$\left.\frac{dW}{d\gamma}\right|_{\gamma=0} = \int_{\Sigma^d} \sum_{i=1}^{n} \left(\int_0^1 u(\theta - \tau(s_{i-1}, s_i, \sigma^d)) - u(\theta - \tau(s_i, s_{i+1}, \sigma^d)) \, dH(\theta|s_i, \sigma^d) \right)$$

$$prob(s_i|\sigma^d)\frac{d s_i}{d\gamma} \, dProb(\sigma^d)$$

$$+ \int_{\Sigma^d} \sum_{i=1}^{n} \int_{s_{i-1}}^{s_i} \int_0^1 -u'(\theta - \tau(s_{i-1}, s_i, \sigma^d)) \left[\frac{\partial\tau(s_{i-1}, s_i, \sigma^d)}{\partial s_{i-1}} \frac{d s_{i-1}}{d\gamma} \right.$$

$$\left. + \frac{\partial\tau(s_{i-1}, s_i, \sigma^d)}{\partial s_i} \frac{d s_i}{d\gamma} \right] dH(\theta|\sigma^p, \sigma^d) \, prob(\sigma^p|\sigma^d)d\sigma^p \, dProb(\sigma^d)$$

$$= 0 + \int_{\Sigma^d} \sum_{i=1}^{n} \left[\frac{\partial\tau(s_{i-1}, s_i, \sigma^d)}{\partial s_{i-1}} \frac{d s_{i-1}}{d\gamma} + \frac{\partial\tau(s_{i-1}, s_i, \sigma^d)}{\partial s_i} \frac{d s_i}{d\gamma} \right]$$

$$\int_{s_{i-1}}^{s_i} \int_0^1 c'(\tau(s_{i-1}, s_i, \sigma^d))dH(\theta|\sigma^p, \sigma^d) \, prob(\sigma^p|\sigma^d)d\sigma^p \, dProb(\sigma^d)$$

$$> 0$$

where the $v()$ terms which cancel out are immediately left out. The second equality holds because of the indifference condition, i.e. a patient with signal s_i is indifferent between reporting the messages (s_{i-1}, s_i) and (s_i, s_{i+1}), and because of the cost sensitive doctor's first order condition for choosing τ. The inequality holds as each s_i is increasing in γ (see lemma 2) and as τ is increasing in s_{i-1} and s_i.

Hence, welfare is increased if copayments are increased from 0.

One could think of a more flexible copayment schedule, i.e. γ could be a function of τ which allows different copayment rates for different treatments. Note that the arguments above are also valid on subsets of the range of possible treatments. Using this insight, proposition 3 is actually more general: There cannot be an interval of treatments prescribed with positive probability such that the optimal copayment rate is zero for the treatments in this interval.

Second, I want to turn to the case where the doctor has no cost incentives. It will be assumed that such a doctor maximizes expected consumer surplus $u(\theta - \tau) - \gamma c(\tau)$ with his treatment decision. Note that this implies that $d\tau(\sigma^p, \sigma^d)/d\gamma < 0$. Again it is shown that $dW/d\gamma$ is positive at $\gamma = 0$. Welfare

in this case is

$$W = \int_\Sigma \int_\Theta v(w-p-\gamma c(\tau(\sigma^p,\sigma^d)))+u(\theta-\tau(\sigma^p,\sigma^d))\ dH(\theta|\sigma^p,\sigma^d)\ d\,Prob(\sigma^p,\sigma^d).$$

Now the derivative of welfare with respect to γ at $\gamma = 0$ is

$$\left.\frac{dW}{d\gamma}\right|_{\gamma=0} = \int_\Sigma \int_\Theta -v'(w-p)\left(c(\tau(\sigma^p,\sigma^d)) + \frac{\partial p}{\partial\gamma}\right) + \frac{d\tau(\sigma^p,\sigma^d)}{d\gamma}$$
$$\left(-u'(\theta-\tau(\sigma^p,\sigma^d)) - v'(w-p)\frac{dp}{d\tau(\sigma^p,\sigma^d)}\right)\ dH(\theta|\sigma^p,\sigma^d)\ d\,Prob(\sigma^p,\sigma^d).$$

The first term is zero by the assumption that insurance profits are zero. The expectation of $u'(\cdot)$ over θ is zero by the first order condition for the doctor's treatment decision. From the zero profit constraint, it is clear that $dp/d\tau(\sigma^p,\sigma^d) > 0$. As $d\tau(\sigma^p,\sigma^d)/d\gamma < 0$, $dW/d\gamma > 0$ at $\gamma = 0$ as had to be shown.

References

Agastya, M., P. Bag, and I. Chakraborty (2012). Communication and authority with a partially-informed expert. working paper available at SSRN.

Armour, B., M. Pitts, R. Maclean, C. Cangialose, M. Kishel, H. Imai, and J. Etchason (2001). The effect of explicit financial incentives on physician behavior. *Archives of Internal Medicine 161*(10), 1261.

Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review 53*(5), 941–973.

Bhattacharya, R., T. Devinney, and M. Pillutla (1998). A formal model of trust based on outcomes. *Academy of Management Review 23*(3), 459–472.

Brook, R. (2010). Physician compensation, cost, and quality. *JAMA: Journal of the American Medical Association 304*(7), 795–796.

Calcott, P. (1999). Demand inducement as cheap talk. *Health Economics 8*(8), 721–733.

Chen, Y. (2009). Communication with two-sided asymmetric information. Technical report, Arizona State University.

Crawford, V. and J. Sobel (1982). Strategic information transmission. *Econometrica 50*(6), 1431–1451.

de Barreda, I. (2010). Cheap talk with two-sided private information. London School of Economics; mimeo.

De Jaegher, K. and M. Jegers (2001). The physician–patient relationship as a game of strategic information transmission. *Health Economics 10*(7), 651–668.

Emanuel, E. and N. Dubler (1995). Preserving the physician-patient relationship in the era of managed care. *JAMA: Journal of the American Medical Association 273* (4), 323–329.

Gallagher, T. and W. Levinson (2004). A prescription for protecting the doctor-patient relationship. *American Journal of Managed Care 10* (2; part 1), 61–68.

Gallagher, T., R. St Peter, M. Chesney, and B. Lo (2001). Patients attitudes toward cost control bonuses for managed care physicians. *Health Affairs 20* (2), 186–192.

Gilson, L. (2003). Trust and the development of health care as a social institution. *Social Science and Medicine 56* (7), 1453–1468.

Glied, S. (2000). Managed care. Volume 1, chapter 13 of *Handbook of Health Economics*, pp. 707–753. Elsevier.

Ishida, J. and T. Shimizu (2010). Cheap talk with an informed receiver. *Institute of Social and Economic Research, Osaka University, Discussion Paper* (746).

Kao, A., D. Green, A. Zaslavsky, J. Koplan, and P. Cleary (1998). The relationship between method of physician payment and patient trust. *JAMA: Journal of the American Medical Association 280* (19), 1708–1714.

Kerr, E., R. Hays, B. Mittman, A. Siu, B. Leake, and R. Brook (1997). Primary care physicians' satisfaction with quality of care in california capitated medical groups. *JAMA: Journal of the American Medical Association 278* (4), 308–312.

Ma, C. and T. McGuire (1997). Optimal health insurance and provider payment. *American Economic Review 87* (4), 685–704.

McGuire, T. (2000). Physician agency. Volume 1, chapter 9 of *Handbook of Health Economics*, pp. 461–536. Elsevier.

Mechanic, D. and M. Schlesinger (1996). The impact of managed care on patients' trust in medical care and their physicians. *JAMA: Journal of the American Medical Association 275*(21), 1693–1697.

Pauly, M. (1968). The economics of moral hazard: comment. *American Economic Review 58*(3), 531–537.

Pitchik, C. and A. Schotter (1987). Honesty in a model of strategic information transmission. *American Economic Review 77*(7), 1032–1036.

Rodwin, M. (1995). Conflicts in managed care. *New England Journal of Medicine 332*(9), 604–607.

Stewart, M. (1995). Effective physician-patient communication and health outcomes: a review. *CMAJ: Canadian Medical Association Journal 152*(9), 1423–1433.