Competition in health care markets: treatment volume, quality and welfare*

Jan Boone†

November 17, 2018

Abstract

This paper introduces a workhorse model to analyze the effects of provider and insurer competition in health care markets. The two contracting imperfections we focus on are the following: (i) whether or not a patient should be treated and (ii) treatment quality are both not contractible. We derive conditions under which the market can implement first best quality and volume with the optimal competition intensities. First best competition intensity is strictly positive in both markets. If there is under-investment in quality, provider competition should be increased. Increasing insurer competition tends to increase treatment volume. If the planner cannot make the provider market competitive enough, it is optimal to increase insurer competition beyond its first best level thereby creating over-treatment.

Keywords: competition in health care markets, insurer competition, provider competition, treatment volume, treatment quality

JEL classification: I13, I11

*Comments and suggestions from seminar participants at Tilec and Tranzo are much appreciated.

†CentER, Department of Economics, Tilec, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands and CEPR, E-mail: *j.boone@uvt.nl*.

1. Introduction

This paper introduces an industrial organization model to analyze the interaction between three health care markets: the health insurance, health provider and health procurement markets. In particular, we focus on the welfare effects of insurer and provider competition. Policy questions we are interested in include: if providers are seen to under-invest in quality, should provider competition be increased or decreased? Or should insurance competition be changed? Can there be too much competition in health markets? Does provider competition cause over-treatment?

Total welfare in health care markets is affected by two fundamental aspects of treatments: quantity and quality. From these follow utility, costs and prices. To illustrate, large scale over-treatment increases costs and hence health insurance premia. Low quality tends to reduce costs but also the value of treatments. So our first questions is: can the market generate the optimal treatment volume and provider investments in quality? We focus on two contractual imperfections. First, a planner (health care regulator) cannot contract on patients' conditions to the extent that first best treatment decisions would become implementable, say through protocols. Providers have an information advantage vis-a-vis both the regulator and insurers. Second, there are treatment quality aspects that are not contractible.

The context we have in mind is the following. There is a health care market with profit maximizing insurers and providers and a health care regulator stipulating (minimum) quality standards. Whereas many countries used to organize their health care sector as a public service, over the years more and more countries introduce competition elements. One rationale given for this is that there are quality elements to health care that are hard to contract by a ministry of health. But patients –it is argued– do experience these quality differences and can inform each other about these. This then gives an incentive for (competing) providers to raise their quality in a way that would not happen under a public system. Improvements in quality due to the introduction of competition elements (in an otherwise public system like the NHS) are, for instance, documented by Propper et al. (2007). See also Gaynor et al. (2015) and OECD (2015) for a discussion of beneficial effects of provider competition in countries like Germany, Sweden, US, Switzerland and the Netherlands. In a similar vein, insurer competition is expected to make insurers more sensitive to consumer wishes offering access to better quality care at lower prices. Porter and Teisberg (2006) and Dafny and Lee (2016) argue that more competition in insurer and provider markets will raise welfare.

In contrast, it is sometimes claimed that competition distorts incentives in health care markets. To illustrate, competing insurers may tend to reduce quality in order to save on costs and lower their premium. If these quality aspects are not contractible, a regulator may not be in a position to prevent this quality downgrade. Moreover, as the quality we focus on is non-contractible, it tends to be non-excludable as well. Thus higher quality induced by one insurer will benefit patients from another insurer too; further reducing insurers' incentives to raise quality. Next, competing providers may treat everyone who walks into their office as they earn their income by treating people; not by pointing out to people that for them treatment is not (cost) effective (Deyo and Patrick, 2005). From this reasoning one could conclude that competition in health care markets leads to under-investment in quality (insurers competing on costs) and over-treatment (providers treating customers to earn an income).

Our analysis shows that the statements above are partial. To illustrate, the over-treatment result is well known in case insurers pay a fee-for-service per treatment where the fee exceeds marginal costs to cover some fixed cost. However, insurers can use two-part-tariffs to overcome this problem: a fee-for-service to cover marginal costs and a capitation fee to cover fixed costs.[1] Further, insurers clearly affect providers' incentives but (non-contractible) treatment quality is set by providers not by insurers.

This paper introduces a workhorse model to analyze the welfare effects of competition via treatment quantity and quality. The model is tractable enough to derive welfare implications of competition intensities in provider and insurer markets.

Our main findings are the following. As a benchmark, we derive two conditions under which the market can implement first best welfare. It turns out that maximizing welfare requires strictly positive competition intensities in both the provider and insurance markets. In this sense, a purely public system without competitive elements is not optimal. To implement first best quality and volume, consumers need to be sufficiently rational/savvy to make the markets competitive enough. If we are not at first best competition intensities, there are four possible cases featuring: over/under-treatment and over/under-investment in quality.

As an illustration of the results, consider the case where patients find the provider market so complicated/in-transparent that first best provider competition cannot be implemented.[2] A planner would then increase insurer competition (even beyond its first best level) creating an outcome with over-treatment and over-investment in quality. If both the provider and insurance markets cannot be competitive enough, there will be over-treatment and under-investment in quality.

We derive policy implications of the following form. If there is under-investment by providers, a regulator can raise provider competition to increase welfare. Under-treatment tends to call for increased insurer competition.

The intuition for our results is as follows. As insurers cannot observe the value and costs of treatment for a particular patient, providers earn an information rent on a treated patient. Providers compete for a patient –to earn this rent– by investing in treatment quality. Ceteris paribus, providers invest more if either the information rent is higher or the provider market is more competitive. If an insurer wants to induce higher quality, the information rent can be raised by increasing the fee-for-service paid for treatment.

We identify two effects of insurer competition intensity on treatment volume. First, there is the insurer competition effect which tends to increase treatment volume ("more competition leads to higher output"). The second effect is driven by insurer's private evaluation of provider quality. If provider competition is so low that insurers' profits are increasing in quality, more intense insurer competition induces insurers to raise treatment volume by increasing the fee-for-service, thereby stimulating quality investments. This strengthens the competition effect. If, in contrast, insurers' profits fall with quality, then in response to more intense insurer competition treatment volume is reduced counteracting the competition effect.

As provider quality is not excludable, first best quality is over-investment from an insurer's point of view. Indeed, quality benefits an insurer's competitor as well. Insurers would like to reduce treatment volume to decrease this (private) over-investment. Insurer competition intensity needs to be sufficiently high to prevent this reduction in treatment volume to sustain

[1] Or other non-linear contracts –like a budget– that can equivalently be implemented using two-part tariffs.

[2] Note that we focus on competition in consumer markets. This is different from a medical-arms-race argument related to competition in the physician labour market: to attract the best physicians, hospitals invest in the latest technologies. Here we focus on hospitals investing in quality to attract patients.

first best quality.

As is known at least since Arrow (1963), there are numerous imperfections in health care markets: asymmetric information leading to moral hazard, adverse selection and contracting problems. We focus on two particular contracting problems.[3] First, the condition of a patient and hence the value of treatment is not verifiable for the insurer and regulator. Hence, the provider decides whether to treat or not. Wennberg (2010, pp. 10) refers to this as "supply sensitive care".[4] This implies that prices will affect providers' treatment decisions. Although whether or not a particular patient should be treated is not contractible, we assume that treatment volume (actual number of patients treated) is. That is, an insurer can pay a fee-for-service for each treated patient. But insurer payment cannot depend on the value of treatment for the patient.

Second, we focus on a quality dimension that is not contractible. Increasing treatment quality is an important motivation for governments to introduce competition in health care markets: see, for instance, Dafny and Lee (2016). Also, van de Ven and Schut (2008), OECD (2015) offer a discussion of the Dutch reforms in 2006 to introduce competition and the motivation for these reforms. Intuitively, if quality would be (completely) contractible, it can be stipulated by a regulator both in a public and a private system. There would be no need to introduce competition to boost this contractible quality.

For concreteness, consider the following examples of non-contractible quality in a health care context. First, the way physicians and nurses communicate with patients. Is this polite, respectful and compassionate? Or are people told they only have to live for one month and are then asked to leave as the next patient is waiting? A hospital can have a protocol that states that each x-ray image of a patient used by a specialist is checked by a separate team the next day. Such a procedure to avoid that things are overlooked only works if it is always done and done with care. But this is hard to verify for an outsider. Porter and Teisberg (2006, pp. 126, 146) argue that a major innovation to increase provider quality is the introduction of a systematic way to measure and record results from treatments and the conditions under which treatments were successful. Another example mentioned by Porter and Teisberg (2006, pp. 51, 373) is regular communication in a hospital between teams/physicians with different specialties. These quality dimensions are not (completely) contractible but patients may be able to notice differences between providers. This will affect their provider choice. A consequence of quality not being contractible is that insurers cannot make it exclusive either; e.g. by claiming that this quality should be applied to their customers but not to the patients of competing insurers.

We do not claim that treatment value or quality is not at all contractible. To illustrate, there do exist protocols specifying under which circumstances a patient should receive a certain treatment (e.g. chemo therapy). Further, with the information recording example: whether the infrastructure for such data gathering is available may be contractible. However, not all treatments come with a verifiable protocol. And the medical staff's efforts to register patient

[3]To keep the model tractable, we ignore adverse selection problems in the health insurance market. Adverse selection in competitive markets can be addressed using risk adjustment. For instance, van de Ven and Schut (2011) argue that the Dutch risk adjustment system has almost eliminated selection incentives for insurers. We come back to this in section 8.

[4]In other words, we focus on supply side moral hazard. With risk averse agents (or agents facing budget constraints) it is not optimal to use demand side cost sharing to fully resolve this type of moral hazard. To simplify the model, we ignore demand side cost sharing.

and outcome information is not contractible. In other words, we focus on the part of these quality dimensions that is not contractible. We are interested here in provider decisions that are not contractible: contractible choices can be set by a regulator.

To the best of our knowledge, this is the first paper that models the interaction in three health markets in order to derive optimal competition intensities for health insurers and health providers. A recent survey of industrial organization aspects in health care markets is Gaynor et al. (2015). They distinguish five stages of interaction between insurers and providers which correspond to the stages that we use in section 2.6 below. Their focus is on empirical findings relating competition to observable outcomes (like mortality) not on the normative results that we derive below (i.e. what is the welfare maximizing competition level). Also, the papers they survey tend to focus on the effects of competition on either a quality measure or price/volume, not their simultaneous determination by competition intensities in both insurer and provider markets. On the other hand, we simplify by focusing on symmetric equilibria where each insurer contracts with each provider. In other words, the analysis below does not address network formation which is covered by Gaynor et al. (2015) and the papers they discuss.

A difference between our paper and the literature on network formation is the formalization of the competition concept. In papers like Ho (2006), Capps et al. (2003) competition and consumer welfare are affected by an insurer's decision on which providers to contract. The number of contracted providers then measures competition intensity. This is also the case for papers analyzing the effects of provider mergers (see, for instance, Dafny, 2009, Ho, 2009, Gowrisankaran et al., 2015). These papers tend to find that a reduction in the number of providers is bad for (consumer) welfare as it increases provider market power. In a similar vein, Ho and Lee (2017) analyze the effects of a reduction in the number of insurers. They find that consumer welfare decreases as the number of insurers falls. Channels through which welfare is affected in these papers by a reduction in the number of insurers/providers include: (i) reduced choice for consumers as the number of suppliers decreases, (iia) more provider market power leads to higher prices paid by insurers/payers, (iib) more insurer market power leads to higher insurer premia paid by consumers, part of which translate –via bargaining– into higher provider prices and (iii) in case of Ho and Lee (2017) the latter effect is not compensated by increased insurer bargaining power vis-a-vis providers.

In contrast, we focus on the aggressiveness of interaction between a given number of providers and insurers on the health provider and insurer markets resp. We argue that there is an optimal competition intensity on each market. Increasing competition further tends to reduce welfare by either causing over-treatment or over-investment in quality.

Whereas the papers mentioned above are empirical in nature, Gal-Or (1997) takes a theoretical approach to network formation. The relation with our work is that this paper analyzes the equilibrium network as a function of the intensity of competition in the health insurer and provider markets. Gal-Or (1997) measures competition as the inverse of the degree of product differentiation (travel costs on a "Hotelling beach"). Our paper keeps the network fixed and analyzes the effect of competition on health care quality and volume.

This paper is organized as follows. The next section introduces the model and characterizes first best treatment quality and volume. Then we consider the incentives facing providers and insurers to invest in quality, set prices and treat patients. Section 4 derives the equilibrium outcome. We illustrate the main results with an example based on Hotelling competition. Section 6 derives conditions under which the market can implement the first best outcome by setting the optimal competition intensities in the health provider and insurer markets. Section 7 shows how the government can determine whether competition intensity in a market should be increased or decreased. We discuss policy instruments for the government to affect the competition intensities. We conclude with a discussion of the policy implications.

2. Model

This section describes the following elements of the model. We start with characterizing the patient's utility from treatment. Then we discuss providers' costs of treatment and quality investment and the contracts offered by insurers on both the health provider and the health insurance markets. First best treatment volume and quality are derived and we explain how we think of competition in this health care context. Finally, we specify the timing of the game that is played out over three markets: the health procurement, the health insurance and the health provider markets.

2.1. patient

We assume that there is one consumer who with probability $\phi \in \langle 0, 1]$ needs one treatment by a provider. We focus on the contracting between insurers and providers and keep the health insurance market as simple as possible. On the demand side we ignore moral hazard, risk aversion and adverse selection. Hence, there is only one type of consumer and ϕ is exogenous. In our model, the treatment decision is taken by the provider. In this sense, there is no need for demand side cost-sharing.[5]

A patient can only get treatment if he has insurance. That is, we assume that the consumer buys health insurance in order to afford treatment. In other words, we focus on access to care as the motivation for buying insurance (Nyman, 1999) not risk aversion. It is well documented that people without health insurance tend to forgo treatment as they have difficulty financing it. These access issues have been stressed both in the popular press (Cohn, 2007) and in academic journals (Nyman, 1999, Schoen et al., 2008, 2010). Many governments are concerned about health consumption inequality caused by income differences and design policies to make health care accessible to low income families (Schokkaert and van de Voorde, 2011). To illustrate this point, president Obama coined his health care reform the *Affordable* Care Act (McDonough, 2011). This is the policy concern we model here. Since we do not focus on risk aversion, we can simplify notation by normalizing $\phi = 1$.

When a patient falls ill, the value of treatment for this patient $v \in [0, \bar{v}]$ is a random draw from a cumulative distribution [density] function $F(v)[f(v)]$. This denotes the value of treatment for the patient in his particular condition. This value is independent from the provider chosen by the patient.

As we derive below, a patient with insurance from insurer I^a and value draw v is treated by provider P^1 if and only if $v < v^a$ where the cut off v^a is determined by the contract between P^1 and the patient's insurer I^a.

[5] As demand side cost sharing cannot fully resolve moral hazard problems, this is without loss of generality. Put differently, we consider the remaining moral hazard problems after demand side cost sharing has been chosen and we ignore possible interaction effects.

Next to treatment volume (captured here by probability of treatment $F(v^a)$), a provider decides on treatment quality, denoted by e. Let provider P^1's effort to raise quality be e^1, then total treatment value to the patient equals $v + e^1$.

The expected value for a patient with insurer I^a to visit P^1 is given by

$$V^{a1} = \int_0^{v^a} (e^1 + v) f(v) dv \tag{1}$$

with similar expressions for V^{b1}, V^{a2}, V^{b2}.[6]

If, instead of using the access to care motive for insurance, we would assume risk aversion, the model would become more elaborate without generating additional insight. To illustrate, we would need to model the provider market for uninsured as well as insured care. But also in this case an expression like (1) would affect from which insurer a customer would buy insurance: in this sense, results would be similar.

2.2. providers

We assume that there are two (ex ante) symmetric providers, denoted P^1, P^2. Providers treat patients and the cost of treatment has two components. First, investing effort e in treatment quality involves a fixed cost $\gamma(e)$ with $\gamma(0) = \gamma'(0) = 0$, and for $e > 0 : \gamma'(e), \gamma''(e) > 0, \lim_{e \to +\infty} \gamma'(e) = +\infty$. This is motivated by Gaynor et al. (2015, pp. 248)'s observation that being "better requires investments in organization and management processes that carry with them substantial fixed costs". One can think here of providers training their staff. Learning better surgery techniques, improving treatment protocols, recording data on which treatments work under which circumstances, reorganize providers along patient needs instead of academic specialties –see, for instance, Porter and Teisberg (2006). These investments are made (largely) independently from the number of treatments provided.[7] Finally, if P^1 invests effort e^1, the quality improvement is enjoyed by customers from both insurers. That is, I^a cannot exclude I^b's patients from benefiting from this quality e^1. If a provider's staff has been trained, they cannot forget their skills when treating a patient from the other insurer.[8]

Second, there is treatment cost $c(v)$ which depends on the patient's condition. We choose the function $c(.)$ so as to capture two aspects: (i) higher fee-for-service leads to more treatments and (ii) setting the fee-for-service too high leads to over-treatment in the sense that treatment value no longer exceeds treatment cost.

[6] As we focus on symmetric equilibria, we can allow for a positive co-payment which is the same for all hospitals and hence drops out of the expressions for provider choice below. To save on notation, we normalize this co-payment to 0.

[7] For ease of exposition, we ignore here two other ways to increase quality. First, one way to improve treatments is to do more of them through learning by doing. Another way to increase treatment value is to use higher quality –and thus more expensive– inputs. Arguably, these two ways are contractible at least to some extent. Learning by doing depends on volume, which we assume can be contracted. Further, in many cases, inputs are contractible as well: e.g. whether titanium or polyethylene is used in a hip replacement.

[8] Some quality aspects can be contracted upon and hence differ between insurers. E.g. whether a patient has a room for his own or shared with other patients. In our model, such differences in quality can be captured by differences in p: the probability of treatment and hence the patient's expected value of treatment can be higher for one insurer than for the other in the same hospital. As we focus on homogeneous consumers and symmetric equilibria, such differences do not materialize in equilibrium.

We make the following assumptions to get these results: $c(0) = c'(0) = 0$ and for $v > 0$ we have $c'(v), c''(v) > 0, \lim_{v \to +\infty} c'(v) = +\infty$: a treatment with value 0 is like having no treatment (no value and no cost); as the treatment value v increases for the patient, it is more costly for the provider. One can think of the patient's condition being more complicated. Treatment then creates higher patient value at a higher cost; say, due to more intense nursing or the operation lasts longer for more complicated cases. This formalization of treatment value and costs has two implications: one positive, one normative.

As we will see, the positive implication is that profit maximizing providers treat more patients if the fee-for-service paid by the insurer is higher. This is in line with empirical findings as summarized in, for instance, McGuire (2011). If the fee-for-service is too low, it becomes unprofitable to treat certain patients although such treatment is valuable from a social point of view. This is sometimes referred to as patient dumping: leaving a patient untreated or sending her/him away from the hospital too early for financial reasons.[9]

This is related to the normative implication. Because $c'' > 0$, for high values of v the condition of the patient has deteriorated so much that the social value of treatment is lower than the cost: $v + e - c(v) < 0$. Hence the set up allows for both (socially) efficient $(v + e - c(v) > 0)$ and inefficient $(v + e - c(v) < 0)$ treatments. This captures public discussions about which treatments should be covered by health insurance. Books like Drummond et al. (2005) and Gold et al. (1996) explain how treatment value can be expressed in quality adjusted life years ("qaly's") and this value should then be compared to the cost of treatment. This corresponds to the comparison of $v + e$ and $c(v)$ above.

A provider decides both how much effort e to invest ex ante (treatment quality) and which patients v to treat ex post (volume). Both decisions are affected by the contracts offered by insurers and the competition intensity on the health provider market. We assume that neither a patient's value v nor the effort investment e is contractible by either an insurer or a regulator.

2.3. insurers

There are two (ex ante) symmetric insurers, denoted I^a, I^b who offer the consumer insurance at a premium σ^a, σ^b resp. Further, they offer providers a contract specifying a two-part tariff. The variable part is the fee-for-service denoted by p. The insurer pays this to the provider in case she treats a patient. Although the patient's treatment value v is not contractible, whether or not a patient is treated is assumed to be contractible. The fixed part T is referred to as either a capitation fee or a prospective payment which is paid irrespective of whether the provider treats a patient. See McGuire (2011) for an overview of supply-side cost sharing contracts used and their effects on health care expenditures.

If insurers would only use linear contracts ($p > 0$ and $T = 0$), we would get similar comparative static results as below. But insurers would be more reluctant to raise p (and thereby treatment volume and quality investments) as they have no second instrument (T) to redistribute rents.

2.4. first best

Suppose that a social planner maximizing total welfare could choose the values for cut off v and effort e; what would the planner choose? There are 2 providers in the market and we assume

[9] See, for instance, Terp et al. (2017) for examples and penalties in the US.

the social planner implements a symmetric outcome –that is, the planner does not decide to close down a provider. Indeed, in the market outcome below, provider competition is needed to get positive quality investments. Hence, we compare the market outcome with the first best outcome featuring two providers. In other words, we focus here on symmetric provider equilibria and take the number of providers as given. The planner chooses v, e to maximize welfare, given by

$$W = \int_0^v (\tilde{v} + e - c(\tilde{v}))f(\tilde{v})d\tilde{v} - 2\gamma(e) \tag{2}$$

A higher value of the cut off v implies that more patients $\tilde{v}$ are treated, creating value $\tilde{v} + e - c(\tilde{v})$.The quality investment $\gamma(e)$ is made by both providers. The first order conditions can be written as

$$\frac{\partial W}{\partial v} = f(v^*)(v^* + e^* - c(v^*)) = 0 \tag{3}$$

$$\frac{\partial W}{\partial e} = F(v^*) - 2\gamma'(e^*) = 0 \tag{4}$$

where v^*, e^* denote first best treatment and quality decisions. Patients v should get treatment as long as the benefit of treatment $v + e$ exceeds the cost $c(v)$. We say that there is over-treatment if a patient v with $v + e < c(v)$ is treated and under-treatment if a patient with $v + e > c(v)$ is not treated.

Increasing e has a marginal benefit equal to 1 for each type v that is treated (treatment happens with probability $F(v^*)$).

2.5. competition

We model competition on the health insurance and health provider markets in similar ways. First, consider the health insurance market. With a probability $\theta_I \in [0, 1]$ insurers compete on this market in the sense that the consumer's choice is affected by the comparison of values offered by each insurer. The higher the value offered by I^a, the more likely it is that the consumer chooses I^a's insurance contract. With probability $1 - \theta_I$ the consumer chooses an insurer randomly. That is, the consumer only checks whether the contract satisfies the individual rationality constraint, but does not check whether the other insurer offers a better deal. In this case –assuming both insurers offer a better deal than the outside option– the probability that the consumer buys from I^a equals $\frac{1}{2}$.

In this simple way, we embed an insurer's monopoly (with probability $\frac{1}{2}(1-\theta_I)$) and competition (with probability θ_I) in one model. Monopoly is defined as the situation where there are two insurers but the consumer chooses his insurer without comparing the value of the contracts on offer.

In case of insurer competition, let V^i denote the consumer value created by insurer $I^i, i = a, b$. Then I^a's market share is given by:

$$\tfrac{1}{2} + \hat{x}^a(V^a - \sigma^a, V^b - \sigma^b) \tag{5}$$

where the twice continuously differentiable function $\hat{x}^a$ has the following properties. First, increasing the value that you offer increases your market share, your opponent's value reduces your market share: $\hat{x}^a_{V^a} > 0, \hat{x}^a_{V^b} < 0$. We assume that the insurance market is fully covered:

$\hat{x}^a(.,.) + \hat{x}^b(.,.) \equiv 0$; that is, the consumer always buys insurance. In the real world, this happens if health insurance is mandatory or if it is attractive enough that everyone buys insurance. Further, we assume symmetry $\hat{x}^a(V,V) = \hat{x}^b(V,V)$ (hence $\hat{x}^a(V,V) = 0$). Market share equals 0 if I^a offers no value, while I^b does ($\hat{x}^a(V^a, V^b) = -\frac{1}{2}$ if $V^a < 0$ while $V^b \geq 0$) and the second derivative satisfies $\hat{x}^a_{V^aV^a} \leq 0$ (to guarantee a concave profit function).

Provider competition is modelled in a similar way. With probability $\theta_P \in [0,1]$ the patient chooses a provider by comparing the values offered by the providers and with probability $1-\theta_P$ he chooses one provider randomly. To illustrate, empirical evidence shows that consumers do respond to hospital quality, but travel distance is a dominant determinant of hospital choice (Pope, 2009.). Hence, a patient can decide to go to the closest provider even if this is not the best one. In a similar vein, Gaynor et al. (2015, pp. 247/8) argue that 50% of heart attack patients arrive in hospital by ambulance; the patient did not choose the hospital in this case. Another rationale for choosing an insurer or provider without comparing values is switching costs. Even though the competitor of your current supplier offers better value, you do not switch (see, for instance, Farrell and Klemperer, 2007, for an overview).

When providers compete, we model this as follows. If the patient bought insurance from I^a (covering both providers) provider P^1 offers value V^{a1} as defined in equation (1) and P^2 offers V^{a2}. As there is no further payment from the patient at this point, provider choice depends on V^{a1}, V^{a2} only. The probability that the patient chooses P^1 –conditional on having I^a insurance– is given by $\frac{1}{2} + \hat{x}^{a1}(V^{a1}, V^{a2})$ where –as with $\hat{x}^a$ above– $\hat{x}^{a1}$ is twice continuously differentiable, $\hat{x}^{a1}(V^{a1}, V^{a2}) + \hat{x}^{a2}(V^{a2}, V^{a1}) = 0$ (full market coverage: as there are no co-payments the consumer always visits a provider), symmetry $\hat{x}^{a1}(V^{a1}, V^{a2}) = \hat{x}^{a2}(V^{a1}, V^{a2})$ (hence $\hat{x}^{a1}(V,V) = 0$) and the derivatives satisfy $\hat{x}^{a1}_{V^{a1}} > 0, \hat{x}^{a1}_{V^{a2}} < 0, \hat{x}^{a1}_{V^{a1}V^{a1}} \leq 0$.

As we have seen, provider value is created by providers investing effort in quality. Although e is not contractible for the insurer, we allow e to affect the patient's choice for a provider.

Broadly speaking, what we have in mind is that patients can learn from each others' experiences by word-of-mouth. Hence, although e is not contractible for the insurer, it can still affect a patient's choice of provider. If a patient is treated well and recovers quickly after treatment, others with the same condition may choose the same provider. This is the standard rationale for a market: aggregating decentralized (quality) information that is not available/contractible at the central level. This may work better for some treatments than for others. For some treatments it may be hard for patients to learn from each others' provider experiences. For such treatments θ_P is low. Such considerations affect the extent to which a provider's market share is affected by her efforts.

We are not claiming that θ_I, θ_P are equal (or close) to one. That is, consumers always comparing the quality aspects of all suppliers. We allow any value between zero and one. Moreover, even if consumers (try to) compare the values offered by the suppliers, they do not necessarily choose the best value with probability 1. We allow $\hat{x}(V', V) < \frac{1}{2}$ for $V' > V$. Indeed, papers like Abaluck and Gruber (2016), Kolstad and Chernew (2009) and Handel and Kolstad (2015) document that consumers can learn but are far from fully rational decision makers. Therefore, in sections 6 and 7 we explicitly consider the cases with low θ as well.

Finally, using θ_I, θ_P to capture/measure competition is in line with the definition of competition intensity in Boone (2000) and Boone (2008). The idea is that a market is called "more competitive" if a firm's payoffs are more sensitive to its efficiency relative to that of competitors. This way of thinking about competition is consistent with well known parametrizations that model increased competition intensity through entry, making goods closer substitutes, reducing travel costs on an Hotelling beach etc. In the context of this paper if $\theta_I(\theta_P) = 0$, insurer (provider) profits are not at all sensitive to the value offered by the other insurer (provider). As $\theta_I(\theta_P)$ increases, an insurer's (provider's) profits are more sensitive to its value compared to the value offered by the competitor. It turns out that with this formalization of competition we find tractable expressions for equilibrium outcomes.

2.6. timing

We analyze a one-shot game with the following timing in three health care markets. In the health procurement market, insurers and providers decide on fee-for-service, capitation fee and quality investments. In the health insurance market, consumers decide from which insurer to buy. Finally, in the health provider market, patients decide which provider to visit and providers decide whether or not to treat the patient.

health procurement market:

stage 1 insurers I^i offer publicly non-discriminatory contracts with a two-part tariff of the form (p^i, T^i) to all providers

stage 2 providers P^j simultaneously and independently decide whether or not to accept each I^i's offer

if provider P^j accepts I^i's contract, she pays I^i the fee T^i; providers and insurers observe the set of accepted contracts

stage 3 providers P^j invest effort e^j at cost $\gamma(e^j)$

health insurance market:

competitive conditions on the insurance market are realized –competition vs. monopoly, with probabilities $\theta_I, 1 - \theta_I$ resp.

stage 4 I^i offers consumer health insurance with a network consisting of the providers who accepted its contract in stage 2 and sets a premium σ^i

stage 5 consumer buys health insurance from at max. one insurer I^i

health provider market:

competitive conditions on the provider market are realized –competition vs. monopoly, with probabilities $\theta_P, 1 - \theta_P$ resp.

stage 6 with probability $\phi = 1$, consumer falls ill and visits P^j –if contracted by patient's insurer I^i

stage 7 provider P^j decides whether or not to treat patient

stage 8 if patient is treated by provider P^j, this provider incurs treatment cost $c(v)$, receives p^i from I^i and the patient derives utility $v + e^j$

Insurer I^i offers each provider the same two-part tariff (p^i, T^i),[10] where p^i denotes the fee-for-service and T^i the capitation fee ($i \in \{a, b\}$). Providers make a profit on treatment –due to the convexity of c– and insurers try to recoup part of this rent via the capitation fee. Hence, we model T^i as a payment from the provider to the insurer.[11] The providers that accept this contract, form I^i's network.

In the first three stages, treatment volume (through p) and treatment quality (through e) are determined before the competition conditions (with probabilities θ_I, θ_P resp.) are realized.

Given the network of providers that have accepted an insurer's contract, the insurer offers health insurance to the consumer on the health insurance market. The consumer buys insurance from one insurer as long as the value of the contract exceeds the consumer's outside option (of no insurance and no treatment) which we normalize to 0.[12]

Then the consumer falls ill and chooses a provider to receive treatment on the health provider market. Finally, the provider who treats the patient gets paid the fee-for-service p^i from the patient's insurer I^i. Patient type v derives utility $v + e^j$ from this treatment by P^j.

3. Incentives for providers and insurers

This section uses backward induction to solve for the equilibrium in the stages of the game in section 2.6. We find the following. A higher fee-for-service set by an insurer, leads to higher provider profits and higher quality investments by providers. Provider competition stimulates quality investments. Competition on the health insurance market tends to increase treatment volume but tends to reduce insurers' incentives to raise quality due to spillover effects.

3.1. health provider market

In the health provider market, the patient decides which provider to visit and the provider decides whether or not to treat the patient. The patient's decision depends on the quality offered by providers and we will analyze this in section 3.3.1 below. Here we characterize the provider's treatment decisions for given quality and prices.

The provider can observe the value v and associated cost $c(v)$ of treatment for a patient, while this is not observable/contractible for the insurer or regulator. If the patient has insurance from I^a, the provider receives p^a per treatment. We assume that providers maximize profits

[10] See Segal (1999) for a discussion of private offers and discriminatory offers. Further, recent papers (see, for instance, Ho and Lee, 2017, Collard-Wexler et al., 2017) assume Nash-in-Nash bargaining on the health procurement market instead of insurers making the offers. This affects the distribution of rents via the capitation fee T. This is relevant in a context where a narrow network can be used by insurers to improve their bargaining position. We, however, focus on symmetric equilibria where each insurer contracts both providers. We leave network formation and alternative bargaining protocols between insurers and providers for future research.

[11] In reality there may be other fixed costs that the insurer reimburses, which we do not model here. T^i then denotes the reduction in this capitation fee paid to the provider. The relevant feature of the capitation fee is not whether it is positive or negative but that it allows the insurer to appropriate at the margin (part of) a change in rents.

[12] Assuming that the competitive situation is revealed before stage 4 simplifies the expression for the premium and insurer profits at this stage. Since we are interested in the insurers' choices at stage 1 –not so much in the premium level– the variables of interest (volume and quality) are determined by taking the expectation over competitive conditions in the insurance market.

and hence are willing to treat patients with $c(v) \leq p^a$.[13] As $c'(v) > 0$, this uniquely defines the cut off value v^a as

$$c(v^a) = p^a \tag{6}$$

and all patient types with $v \leq v^a$ are treated. As we assume that providers are symmetric, they have the same cost function and hence cut-off v^a is a property of insurer I^a's contract. By being more generous –higher p^a– insurer I^a increases the probability that its patient is treated. Hence, we can write I^a's contract as $(p^a, T^a) = (c(v^a), T^a)$. To simplify notation, we write the contracts directly as (v^a, T^a).

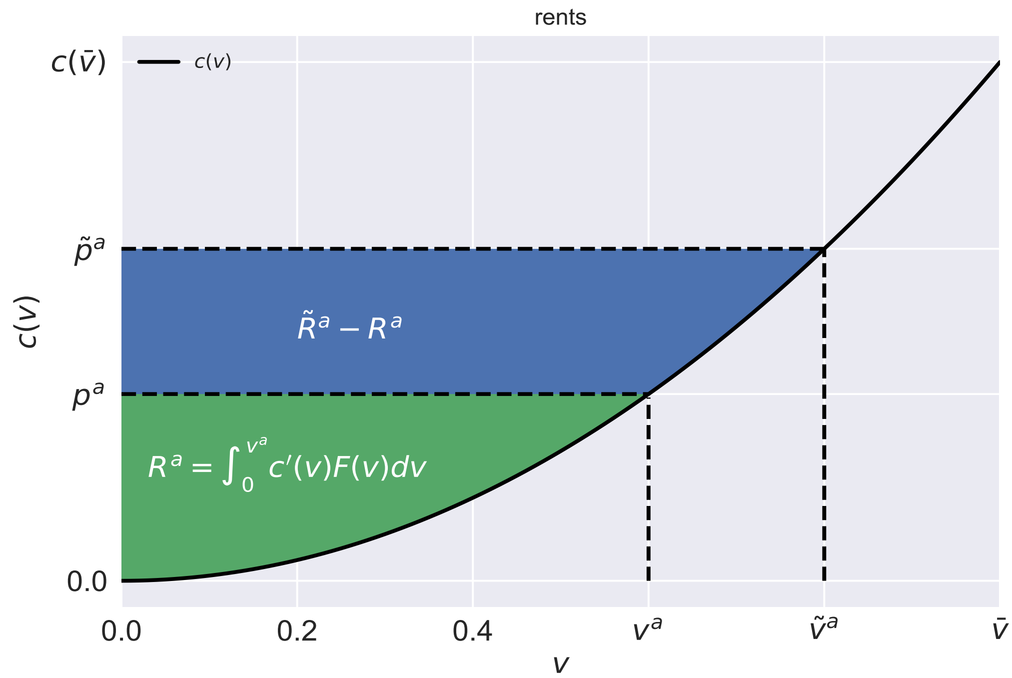

Figure 1: Provider rents as a function of the fee-for-service p.

By the assumption that $c'(v) > 0$, providers earn a profit on the infra-marginal patient types since $p^a > c(v)$ for $v < v^a$. Hence, if a patient with insurance from I^a visits provider P^1, the provider earns the following expected rent on this patient:

$$R^a = \int_0^{v^a} (p^a - c(v))f(v)dv \tag{7}$$

This rent is illustrated in figure 1. The important feature is that as the fee-for-service p increases (from p^a to $\tilde{p}^a$), treatment volume v increases (from v^a to $\tilde{v}^a$) and the providers' rent R increases (from R^a to $\tilde{R}^a$). Using $p^a = c(v^a)$ and integration by parts, it is routine to find the following.

[13] We do not allow providers to commit to treating more (or less) patients than is profitable ex post when the patient visits the provider.

Lemma 1. *The expected rent earned by a provider treating I^a's patient can be written as*

$$R^a = \int_0^{v^a} c'(v)F(v)dv = \int_0^{v^a} c'(v)\frac{F(v)}{f(v)}f(v)dv \tag{8}$$

The second equality shows the similarity with the information rent in a model with asymmetric information. The asymmetric information here is about patient type v. A provider can earn an expected profit R^a if I^a's patient chooses her for treatment.

Note that the rent R^a is caused by asymmetric information, not by (lack of) provider competition. Provider competition intensity determines the effort invested by providers to capture the patient and thereby earn this rent.

3.2. health insurance market

On the health insurance market, assume for now that I^a gives the incentive to both providers to invest e effort and treat all patients with $v < v^a$. Then the value created by I^a for the consumer is given by

$$V^a = \int_0^{v^a} (e + v)f(v)dv \tag{9}$$

A patient gets treated if $v < v^a$, in which case utility equals $e + v$. Without insurance the patient has no access to this treatment and his utility is normalized to 0.

As the competitive situation on the health insurance market is revealed before insurers set their premium, with probability $1 - \theta_I$, the consumer chooses a contract randomly –conditional on it creating non-negative value– and in this case the insurer sets a premium equal to $\sigma^a = V^a$ to make the value of the contract, $V^a - \sigma^a$, equal to the outside option with zero-value. The expected profit for I^a on the (monopoly) health insurance market is then given by

$$\pi^{ma} = \tfrac{1}{2}(V^a - F(v^a)p^a) \tag{10}$$

where the probability of choosing I^a equals $\frac{1}{2}$, p^a is paid to the provider in case the patient is treated which happens with probability $F(v^a)$.

With probability θ_I insurers compete and then I^a's (competitive) profits are given by

$$\hat{\pi}^{ca} = \max_{\mu^a} \mu^a(\tfrac{1}{2} + \hat{x}^a(V^a - \mu^a - F(v^a)p^a, V^b - \mu^b - F(v^b)p^b)) \tag{11}$$

where the margin is defined as the difference between premium and expected fee-for-service $\mu^a = \sigma^a - F(v^a)p^a$ and the function $\hat{x}^a$ was introduced in section 2.5.

Let π^{ca} denote the Nash equilibrium profits after solving for the Nash equilibrium in μ^a, μ^b.[14] Then we can show the following.

Lemma 2. *In the competition state on the health insurance market a pure strategy Nash equilibrium exists. The Nash equilibrium profits and market shares can be written as*

$$\pi^{ca}(V^a - F(v^a)p^a, V^b - F(v^b)p^b) \text{ and } x^a(V^a - F(v^a)p^a, V^b - F(v^b)p^b), \tag{12}$$

[14] To be explicit $\hat{x}^a, \hat{\pi}^{ca}$ depend on insurers' choice variables μ^a, μ^b. The functions x^a, π^{ca} denote market share and profits after solving for the equilibrium choices μ^a, μ^b.

the derivative of π^{ca} with respect to its first argument is given by

$$\pi^{ca}_{V^a} = (\tfrac{1}{2} + x^a)(1 - \psi) > 0 \tag{13}$$

where $\psi \in \langle 0, 1 \rangle$.

Equation (12) follows from equation (11) where the optimization depends on terms of the form $V^i - F(v^i)p^i$. By writing the optimization in (11) in terms of μ (instead of σ) makes this immediately clear. To ease the exposition, we assume that the Nash equilibrium in μ^a, μ^b is unique.

The intuition for equation (13) is as follows. As V^a increases, I^a can increase σ^a which raises I^a's profits at the rate of its market share $\frac{1}{2} + x^a$. However, the increase in V^a worsens I^b's competitive position and I^b will react to this by reducing its premium σ^b. With price competition, reaction functions are upward sloping and I^a reacts to the reduction in σ^b by reducing σ^a –that is, increasing σ_a by less than would happen for unchanged σ_b. Hence, I^a's profits increase with less than $(\frac{1}{2} + x^a)dV^a$. The "discount" on this profit increase is given by $1 - \psi$. Thus under competition, an increase in V^a with 1, increases expected utility with 1 for $\frac{1}{2} + x^a$ consumers. But I^a can only appropriate share $1 - \psi < 1$ from this increase in terms of profits. Under monopoly, in contrast, $\pi^{ma}_{V^a} = \frac{1}{2}$: the insurer appropriates the full increase in expected utility –due to an increase in V^a– as an increase in profits.

Expected profits for I^a before the monopoly/competition regime is resolved are given by

$$\begin{aligned}\pi^a &= (1 - \theta_I)\pi^{ma} + \theta_I \pi^{ca} \\ &= \tfrac{1}{2}(1 - \theta_I)(V^a - F(v^a)p^a) + \theta_I \pi^{ca}(V^a - F(v^a)p^a, V^b - F(v^b)p^b)\end{aligned} \tag{14}$$

and I^a's expected market share equals

$$ms^a = \tfrac{1}{2} + \theta_I x^a(V^a - F(v^a)p^a, V^b - F(v^b)p^b) \tag{15}$$

With $\theta_I = 0$ we say that there is insurer monopoly where the insurer is chosen with probability 1/2 independently from the value offered. As θ_I increases, both insurer market share and profits depend more on value offered and we say the market is more competitive.

Overall competition faced by I^a is determined by $\theta_I x^a_{V^a}$, where x^a derives its properties from $\hat{x}^a$ above. Section 5 gives an example to illustrate the relation between x^a and $\hat{x}^a$. Loosely speaking, we think of $x^a_{V^a}$ as determining the "exogenous" competition intensity in the market, while θ_I denotes the part that can be adjusted by a social planner/the government.

3.3. health procurement market

On the health procurement market, providers choose their effort levels given the contracts offered by insurers. The equilibrium provider efforts are determined by the competition regime on the provider market. After characterizing provider effort, we proceed with backward induction to determine insurers' contracts in stage 1.

3.3.1. providers

With probability $1 - \theta_P$ the patient does not consider value differences between providers and randomly chooses a provider to visit; say, the one closest to his home. Using the function $\hat{x}^{a1}$

introduced above, the expected market share for P^1 on I^a's market is given by

$$ms^{a1} = \theta_P(\tfrac{1}{2} + \hat{x}^{a1}(V^{a1}, V^{a2})) + (1-\theta_P)\tfrac{1}{2} = \tfrac{1}{2} + \theta_P \hat{x}^{a1}(V^{a1}, V^{a2}) \tag{16}$$

The stronger this market share depends on values created, the more competitive we call the provider market. The dependency of P^1's market share on V^{a1} is given by $\theta_P \hat{x}^{a1}_{V^{a1}}$. Again we interpret $\hat{x}^{a1}_{V^{a1}}$ as being exogenous to the market and θ_P the parameter that a planner/government can change.

We follow Gal-Or (1997) in assuming that the consumer when buying insurance from I^a does not know yet what his provider preference will be once he needs treatment: insurer choice is driven by V^a. Gal-Or motivates this assumption in her set-up with the idea that providers may offer different types of treatments for a number of conditions (horizontal differentiation) and ex ante the consumer does not know for which condition he will need treatment and which treatment type he then prefers. After falling ill, his condition is revealed and he ends up with a particular preference determining whether or not he chooses P^1 if, say, $V^{a1} > V^{a2}$.

Providers choose their effort e in stage 3 before they know which insurer is chosen by the patient on the health insurance market and before they learn which competition regime (with probabilities θ_P and $1-\theta_P$) is realized on the health provider market.

Above we derived the equilibrium on the health insurance market. Using lemma 2, we write P^1's profit as:

$$\hat{\pi}^1(e^1, e^2) = (\tfrac{1}{2} + \theta_I x^a)[\tfrac{1}{2} + \theta_P \hat{x}^{a1}]R^a + (\tfrac{1}{2} - \theta_I x^a)[\tfrac{1}{2} + \theta_P \hat{x}^{b1}]R^b - \gamma(e^1) - T^a - T^b \tag{17}$$

where we use short-hands like: $x^a = x^a(V^a - F(v^a)c(v^a), V^b - F(v^b)c(v^b))$, $\hat{x}^{a1} = \hat{x}^{a1}(V^{a1}, V^{a2})$.

The first expression in round brackets gives I^a's market share. Conditional on the consumer buying insurance from I^a, the expression in square brackets gives the market share for P^1 (see equation (16)). Hence, the product of the term in round and square brackets gives the joint probability of the consumer choosing I^a and P^1 for treatment. Similarly, the next product of round and square brackets gives the probability of the consumer choosing I^b and P^1.

If the consumer buys from I^a and visits P^1, P^1 earns an expected rent R^a as given by (8); similar expression for I^b and R^b. Further, P^1 incurs the cost of effort $\gamma(e^1)$ and pays the fees T^a, T^b which are sunk when e^1 is chosen.

The following expressions for derivatives are used below:

$$\begin{aligned} x^a_{e^1} &= x^a_{V^a}F(v^a) + x^a_{V^b}F(v^b) && (18)\\ \hat{x}^{a1}_{e^1} &= \hat{x}^{a1}_{V^{a1}}F(v^a) && (19) \end{aligned}$$

As e^1 increases, V^a increases with $F(v^a)$ –the probability that the patient benefits from the increased quality. Hence, I^a's market share increases with $x^a_{V^a}F(v^a)$. However, as e^1 is not exclusive for I^a, V^b increases as well with e^1. This reduces I^a's market share with $x^a_{V^b}F(v^b)$. In other words, e^1 only increases I^a's market share if the patient is more likely to be treated with insurance from I^a ($F(v^a) > F(v^b)$) as he is more likely to benefit from this higher quality than with insurance from I^b. Further, an increase in e^1 increases V^{a1} with $F(v^a)$ and thus P^1's market share on I^a's market with $\hat{x}^{a1}_{V^{a1}}F(v^a) > 0$.

We focus on a symmetric equilibrium in effort levels. The first order condition for e^1, $\hat{\pi}^1_{e^1} = 0$, can be written as

$$\begin{aligned} &R^a[\tfrac{1}{2}\theta_I(x^a_{V^a}F(v^a) + x^a_{V^b}F(v^b)) + (\tfrac{1}{2} + \theta_I x^a)\theta_P \hat{x}^{a1}_{V^{a1}}F(v^a)] \\ &\quad + R^b[-\tfrac{1}{2}\theta_I(x^a_{V^a}F(v^a) + x^a_{V^b}F(v^b)) + (\tfrac{1}{2} - \theta_I x^a)\theta_P \hat{x}^{b1}_{V^{b1}}F(v^b)] - \gamma'(e^1) = 0 \end{aligned} \tag{20}$$

where we use that in a symmetric provider equilibrium with $e^1 = e^2 = e$ we have $\hat{x}^{a1} = \hat{x}^{b1} = 0$ because $V^{a1} = V^{a2}$ and $V^{b1} = V^{b2}$. It is routine to verify that the derivative of (20) with respect to e^1 is negative in case $v^a = v^b$ (see (A.10) in the appendix). Off the equilibrium path, we assume that P^j's objective function is concave in e^j also if $v^a \neq v^b$.

Note that in symmetric insurer equilibrium where we have both $v^a = v^b = v$ and $R^a = R^b = R$, we can simplify (20) to

$$\gamma'(e) = \theta_P \hat{x}^{a1}_{V^{a1}} F(v) R(v) \tag{21}$$

In case of monopoly ($\theta_P = 0$), increasing quality and hence patient value does not affect provider choice (say, because a patient simply goes to the closest hospital irrespective of quality offered). Consequently, provider revenues are not affected and there is no incentive for providers to invest in quality (above the minimum required level). They only invest effort if there is competition ($\theta_P > 0$) on the provider market. Increasing e then raises value at the rate $F(v)$ and thus market share at the speed $\hat{x}^{a1}_{e^1}$. If the patient chooses this provider, she earns $R(v)$ in expected terms.

This is one of the two equations that hold in symmetric equilibrium to characterize v and e. But when I^a changes v^a, it cannot assume that v^b changes as well. Hence, for I^a considering different values for v^a the relevant equation is (20).

This equation (20) is the incentive compatibility (IC) constraint that insurers need to take into account when setting $v^{a,b}$. It specifies the relation between e^1 and v^a. To find the effect of v^a on e^1, we start from (20) and impose that I^a knows that the providers –for each value of v^a– play a symmetric equilibrium with $e^1 = e^2 = e$. We first take the derivative with respect to v^a and (only then) evaluate this derivative in the symmetric equilibrium with $v^a = v^b = v$. We have the following result.

Lemma 3. *A pure strategy Nash equilibrium in e^1, e^2 exists. In a symmetric equilibrium with $e^1 = e^2 = e$, when insurer I^a increases v^a, it expects the following increase in $e^1 = e^2 = e$:*

$$\left.\frac{de}{dv^a}\right|_{e^1=e^2, v^a=v^b} = \frac{1}{2}\frac{\theta_P \hat{x}^{a1}_{V^{a1}} \frac{d(R(v)F(v))}{dv}}{\gamma''(e) - \theta_P R(v) F(v)^2 \hat{x}^{a1}_{V^{a1}V^{a1}}} > 0 \tag{22}$$

where the derivative is evaluated at the symmetric outcome $v^a = v^b = v$.

For $\theta_P > 0$, this derivative is positive: by increasing v^a, I^a increases rents R^a to providers and thus provider effort: it becomes more attractive for the provider to "capture" the patient by offering high quality e. Further, an increase in v^a induces a bigger increase in effort e (i) the more elastic effort is (lower denominator), (ii) the more market share reacts to value offered ($\theta_P \hat{x}^{a1}_{V^{a1}}$) and (iii) the faster $R(v)F(v)$ increases with v^a: $R(v)$ is the profit of gaining a patient and $F(v)$ is the rate at which V^{a1} increases with e^1; as $d(R(v)F(v))/dv$ is higher, the same increase in v leads to a bigger increase in e.

3.3.2. insurers

Moving back to the start of the game. Insurer I^a in stage 1 sets v^a, T^a (and thus fee-for-service $p^a = c(v^a)$) to maximize profits subject to the providers' IC constraint (20). First, consider the level of the capitation fee, T^a. I^a's profits in symmetric equilibrium can be written as

$$\Pi^a = \pi^a + 2 * T^a \tag{23}$$

where π^a is given in equation (14) and I^a receives T^a from both providers (if they accept). As I^a makes the offer, it sets T^a to make providers indifferent between accepting and rejecting the offer. To derive T^a, we write the profits in equation (17) temporarily as $\Pi^1 = \pi^1 - T^a - T^b$. Let $\pi^{1,-a}$ denote P^1's profits in case she rejects I^a's offer and P^2 follows the equilibrium by accepting both insurers' offers. Then we find that

$$\Pi^{1,-a} = \pi^{1,-a} - T^b \tag{24}$$

Making P^1 indifferent between accepting and rejecting (that is, $\Pi^1 = \Pi^{1,-a}$) I^a sets T^a such that

$$T^a = \pi^1 - \pi^{1,-a} \tag{25}$$

In words, the fixed fee equals I^a's contribution to P^1's profits.

We consider an equilibrium where each insurer contracts both providers. To characterize the payments $T^{a,b}$ in this equilibrium, we need to specify out-of-equilibrium behavior in case, say, I^a and P^1 do not have a contract. In particular, in this case there is a probability $(1-\theta_I)(1-\theta_P)/4$ that the consumer has an explicit preference for both I^a and P^1 irrespective of values offered. The proof of the next lemma starts with an assumption resolving this technical issue.

Lemma 4. *Consider an equilibrium where each insurer contracts both providers and where providers play a symmetric equilibrium in stage 3 where $e^1 = e^2 = e$ is determined by (21). In this equilibrium insurer I^a's (expected) profits in stage 1 can be written –up to a constant for I^a– as*

$$\Pi^a = (1-\theta_I)\pi^{ma} + \theta_I \pi^{ca} + (\tfrac{1}{2} + \theta_I x^a)R^a + (\tfrac{1}{2} - \theta_I x^a)R^b - 2\gamma(e) \tag{26}$$

Profits feature the following partial derivatives:

$$\frac{\partial \Pi^a}{\partial v^a} = f(v^a)\left[(e + v^a - c(v^a))\tfrac{1}{2}(1-\psi\theta_I) + \tfrac{1}{2}\psi\theta_I c'(v^a)\frac{F(v^a)}{f(v^a)}\right] \tag{27}$$

and

$$\frac{\partial \Pi^a}{\partial e} = \tfrac{1}{2}(1-\theta_I)F(v^a) - 2\gamma'(e) \tag{28}$$

The intuition for these expressions is as follows. First, Π^a is made up of the expected profits for the insurer on the health insurance market, where the expectation is taken over the competition regime that will prevail in this market; profit levels π^{ma}, π^{ca} are defined in (10, 12). In addition to this, I^a recoups part of the provider rents (R^a, R^b). This rent expression follows from equation (17) with $\hat{x}^{a1} = \hat{x}^{b1} = 0$ (recall that in symmetric provider equilibrium: $e^1 = e^2$) and the rent is collected from two providers. At the margin I^a's profits are affected by R^b. That is, if I^a manages to increase x^a, it can appropriate the increase in the providers' expected rent $\theta_I x^a R^a$ but –at the same time– it needs to compensate the providers if they lose out on R^b. The latter effect then reduces the transfer T^a that I^a can ask from providers. Finally, if I^a wants to induce higher e, it has to compensate both providers for the increase in cost $\gamma(e)$.

We can write the first order condition for v^a as follows

$$\frac{d\Pi^a}{dv^a} = \frac{\partial \Pi^a}{\partial v^a} + \frac{\partial \Pi^a}{\partial e^a}\frac{de}{dv^a} = 0 \tag{29}$$

where de/dv^a is given by the providers' IC constraint (22). To get an intuition for the derivatives $\partial \Pi^a/\partial v^a, \partial \Pi^a/\partial e^a$, first we look at each effect separately (*ceteris paribus*). This can be interpreted as the case where e is contractible for insurers. In this case, the first order conditions for v and e can be written as:

$$f(v)\left[(e+v-c(v))\tfrac{1}{2}(1-\psi\theta_I)+\tfrac{1}{2}\psi\theta_I c'(v)\frac{F(v)}{f(v)}\right]=0 \tag{30}$$

$$\tfrac{1}{2}(1-\theta_I)F(v)-2\gamma'(e)=0 \tag{31}$$

In case of insurer monopoly ($\theta_I = 0$), v in (30) is set at its first best level (for given effort e): $e + v - c(v) = 0$ as in (3). Under monopoly, the insurer sets a premium equal to the consumer surplus and hence has an incentive to set v (and hence $p = c(v)$) to maximize this surplus. Conditional on e, the insurer chooses efficient treatment volume.[15] However, under monopoly e is not chosen optimally. Comparison with first best effort in (4) shows that there is under-investment in treatment quality. Recall that insurer monopoly means that consumers do not choose the insurer based on value offered; there are still two insurers. As the quality investments are not excludable, the effort that is induced by I^a's choice of p^a benefits the patient of I^b as well. This positive externality leads to under-investment in quality. The social planner understands that an increase in e benefits the patient no matter which insurer he chooses; for I^a there is only a benefit (increase in profits) if the patient chooses its contract: $\frac{1}{2}$ in front of $F(v)$ in (31).

With insurer competition $\theta_I > 0$, volume v is increased compared to insurer monopoly. As the second term in in square brackets in (30) is positive, the benefit of treatment volume is increased and competition raises "output".

The intuition for this competition effect is as follows. An increase in v^a has the following effect on value created:

$$\frac{\partial\left(\int_0^{v^a}(e+v)f(v)dv-F(v^a)c(v^a)\right)}{\partial v^a}=\underbrace{f(v^a)(e+v^a-c(v^a))}_{\text{efficiency effect}}-\underbrace{F(v^a)c'(v^a)}_{\text{rent effect}} \tag{32}$$

The efficiency effect is identical to the planner's tradeoff in (3). Next, increasing v, the insurer has to pay a higher fee-for-service $p = c(v)$ also for patients with lower costs. This is the rent effect over the infra-marginal patients $F(v^a)c'(v^a)$. Consider the competition case ($\theta_I = 1$) and start at first best ($e + v - c(v) = 0$), what is the effect on I^a's profits of increasing v^a further? The first order effect on efficiency equals 0. Increasing p raises rents to providers which I^a can recapture –at the margin– via T^a. From (26), this benefit for I^a equals $(\frac{1}{2} + x^a)dR^a/dv^a = (\frac{1}{2} + x^a)c'(v^a)F(v^a)$. Of course, this benefit is offset in the health insurance market as I^a faces higher costs of its insurance contract. As shown in equation (13), the effect on profits of this reduction in value offered $V^a - F(v^a)c(v^a)$ is of the order $(\frac{1}{2} + x^a)(1-\psi)c'(v^a)F(v^a)$. Summing these two effects gives

$$(\tfrac{1}{2}+x^a)\psi c'(v^a)F(v^a)=c'(v^a)F(v^a)\psi/2>0 \tag{33}$$

in symmetric equilibrium with $x^a = 0$.

[15] This efficiency result is not generic. If the monopolist cannot appropriate the whole consumer surplus, it sets treatment volume too low (as in a standard monopoly problem). The generic feature is that more intense insurer competition leads to higher volume, as we see below.

Intuitively, because of price competition in the health insurance market, I^a's cost increase ($c'(v^a)F(v^a)$) is partly ($\psi \in \langle 0, 1\rangle$) translated into an equilibrium price (premium) increase. Hence, I^a captures the full increase in rents with T^a and suffers a loss in profits (due to increased costs) on the health insurance market but the latter is smaller. Part of the cost increase is transferred to consumers by charging a higher equilibrium price. This is not possible under monopoly in our model since I^a then already charges $\sigma^a = V^a$ which cannot be increased further. But with price competition, the increase in costs is partly passed on through a higher premium σ^a. This, in turn, induces higher σ^b (upward sloping reaction functions). By increasing fee-for-service p^a, I^a makes itself a weaker competitor in the health insurance market. This raises profits with upward sloping reaction functions. This is sometimes referred to as a "fat cat" strategy (Bulow et al., 1985, Fudenberg and Tirole, 1984). As insurer competition θ_I intensifies, this pass-through or fat cat effect is stronger, and hence the fee-for-service p^a increases and treatment volume increases, too.

Whereas insurer competition increases treatment quantity, it reduces treatment quality. Equation (31) implies that at $\theta_I = 1$ there is no incentive to invest in quality for insurers. With competition, quality is valuable to the extent that it raises V^a compared to V^b. But since the quality we consider here is not excludable, I^b benefits from the increase in e induced by I^a. This implies that insurer market power ($\theta_I < 1$) is needed to create an incentive to invest in treatment quality. At "full competition" ($\theta_I = 1$) there is no investment in quality, if such investments would be contractible (but not excludable).

However, we assume that quality is not contractible. Although useful for illustration, the relevant first order condition is (29) –not (30,31). By setting v, rents R are created which lead providers to invest in quality if $\theta_P > 0$. Hence, even with $\theta_I = 1$ there will be positive investments in quality.

4. Equilibrium quality and quantity

Combining the equation determining treatment volume (29) with (21) determining equilibrium treatment quality, we characterize the equilibrium values for treatment quantity and quality.

First, define the following three effects: Treatment Efficiency, Direct Competition and Private Marginal Value of Quality. We write these as functions of v only.

$$\text{TE}(v) = e(v) + v - c(v) \tag{34}$$

$$\text{DC}(v) = \frac{\psi}{2}c'(v) \tag{35}$$

$$\text{PMVQ}(v) = \tfrac{1}{2}(1-\theta_I) - 2\theta_P R(v)\hat{x}^{a1}_{V^{a1}} \tag{36}$$

where from (21) equilibrium quality effort is determined by

$$\gamma'(e(v)) = \theta_P R(v)F(v)\hat{x}^{a1}_{V^{a1}} \tag{37}$$

and

$$\hat{x}^{a1}_{V^{a1}} = \hat{x}^{a1}_{V^{a1}}\left(\int_0^v (e(\tilde{v}) + \tilde{v})f(\tilde{v})d\tilde{v}, \int_0^v (e(\tilde{v}) + \tilde{v})f(\tilde{v})d\tilde{v}\right) \tag{38}$$

TE is based on $\partial W/\partial v$ in equation (3). DC is the pass through term in case of competition as derived in (33). Finally, PMVQ is based on $\partial \Pi^a/\partial e$ in equation (28) where we substitute the equilibrium value for $\gamma'(e)$ in (37).

Proposition 1. *If a symmetric equilibrium exists, v is determined by*

$$TE(v)\frac{1-\psi\theta_I}{2} = -\frac{F(v)}{f(v)}\left(\theta_I DC(v) + PMVQ(v)\frac{de}{dv^a}\right) \tag{39}$$

where de/dv^a is given by (22).

In an interior equilibrium, the right hand side of this equation cuts the left hand side from below.

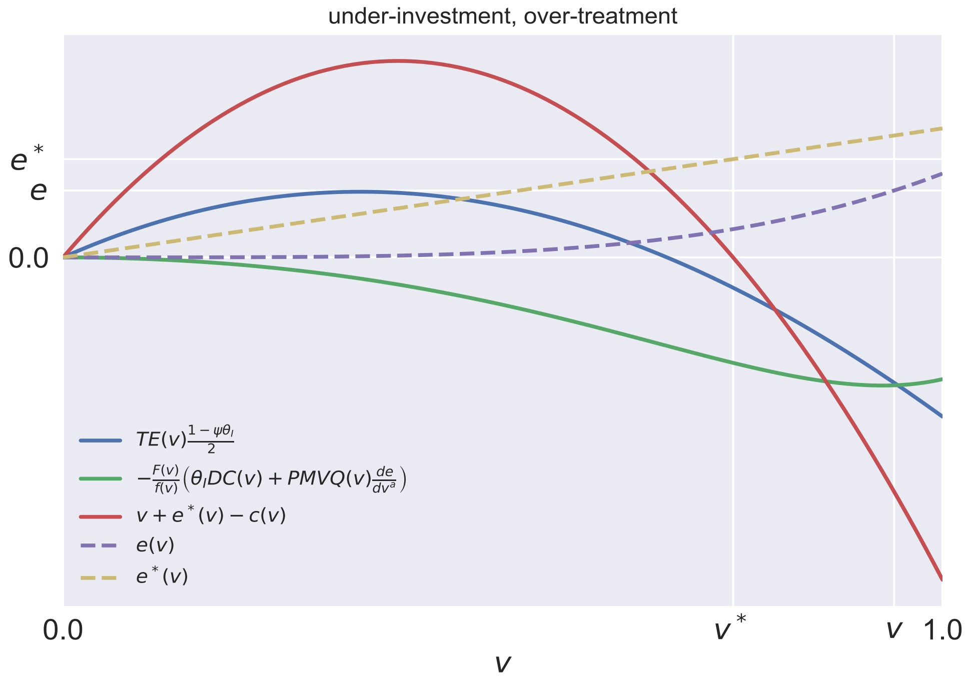

Figure 2: Intersection of the red line (3) with the horizontal axis gives first best v^*; the dashed line $e^*(v)$ then gives first best quality $e^* = e^*(v^*)$. Intersection of the blue (left hand side of (39)) and the green (right hand side) lines gives market equilibrium v; dashed line $e(v)$ gives the corresponding quality level.

A well known result of the literature on health care competition and quality is the following. If prices are exogenous (e.g. set by a regulator) then more intense provider competition tends to lead to higher quality. This effect has been documented for the NHS in the UK and for the Medicare program in the US (see Gaynor et al., 2015, for an overview). If prices are exogenous, then p (and hence v) is fixed and equation (39) is no longer relevant. In this case, equation (37) immediately implies that higher $\theta_P \hat{x}^{a1}_{V^{a1}}$ (more intense provider competition) leads to higher effort e and hence higher quality. Our model is consistent with this empirical finding. If prices are set endogenously, the effect is ambiguous both empirically (Gaynor et al., 2015, pp. 249) and in our model.

Figure 2 illustrates an equilibrium: the intersection of the blue curve (left hand side of (39)) and the green curve (right hand side). The equilibrium industry effort curve $e(v)$ is given by

$$e(v) = (\gamma')^{-1}(\theta_P R(v) F(v) \hat{x}^{a1}_{V^{a1}}) \tag{40}$$

The first best equivalent of this curve is given by

$$e^*(v) = (\gamma')^{-1}(\tfrac{1}{2}F(v)) \tag{41}$$

For the parameter values used in figure 2, we have $\theta_P R(v)\hat{x}^{a1}_{V^{a1}} < \frac{1}{2}$ and $e^*(v)$ lies above $e(v)$ for all $v \in \langle 0, 1]$.

First, if the right hand side of (39) equals 0, we have treatment efficiency, TE = 0. If it is positive (negative, as in figure 2), we have under (over)-treatment.

The DC term has the effect that insurer competition raises treatment volume. This is the *competition effect* of ψ discussed above. As DC increases, the right hand side of (39) –green curve– shifts down, leading to higher equilibrium treatment volume v.

The PMVQ effect is driven by the quality effect on insurer profits; in particular, $-(\partial\Pi^a/\partial e)(de/dv^a)$. As $de/dv^a > 0$ in equation (22), the sign of this term is determined by $-\partial\Pi^a/\partial e$. If PMVQ $= \partial\Pi^a/\partial e < 0$, quality investments reduce I^a's profits at the margin; there is over-investment in quality *from the insurer's point of view*. By decreasing v, I^a reduces e and hence increases its profits. If PMVQ becomes more negative, the right hand side –green curve– shifts upwards, reducing treatment volume v. In contrast, if $\partial\Pi^a/\partial e > 0$, I^a wants to increase v as the consequent increase in e raises profits: there is under-investment in quality *from the insurer's point of view*. The sum of the DC and PMVQ effects determines whether there is over or under-treatment in equilibrium.

An intuitive way to think about the equilibrium is as follows. Provider competition intensity $(\theta_P\hat{x}^{a1}_{V^{a1}})$ determines treatment quality, for given volume. If $\theta_P\hat{x}^{a1}_{V^{a1}}R(v) = \frac{1}{2}$, we have first best quality, given v: compare (37) and (4). Higher (lower) θ_P then leads to over (under) investment in quality. For given e, whether or not treatment volume is efficient depends on the DC versus the PMVQ effects. If there is under-investment from the private point of view, the two effects strengthen each other leading to over-treatment. If there is over-investment from a private point of view, the competition effect is mitigated. For θ_I small enough, there will be under-treatment.

The next lemma gives a condition for the existence of an equilibrium in the game presented here. We have an interior solution for v if the blue and green curves (left and right hand sides of (39)) intersect. If not, we have a corner solution $v = \bar{v}$.

Lemma 5. *Assume that Π^a is concave in v^a, e. Then there exists a symmetric equilibrium with an interior solution for $v \in [0, \bar{v}]$ if the left and right hand side of (39) intersect at $v > 0$. Otherwise, we have $v = \bar{v}$ and $e = (\gamma')^{-1}(\theta_P R(\bar{v})\hat{x}^{a1}_{V^{a1}})$.*

Note that $e = v = 0$ is not an equilibrium outcome in figure 2: by raising p^a, I^a can raise e and v thereby creating an increase in profits.

5. Example

This section uses an Hotelling model to parametrize the functions $\hat{x}^a, \hat{x}^{a1}$ etc. This example helps to illustrate the results above and we use it to plot the figures in this paper.

First, consider the health insurance market which is competitive with probability θ_I. In case of competition, insurers compete on a Hotelling "beach" of length one and there is a uniform distribution of consumer types over the beach; i.e. the probability that the consumer is located between the arbitrary points $x \geq 0$ and $y \in \langle x, 1]$ is given by $y - x$. Insurer I^a is placed on the far left and I^b on the far right. Travel costs per unit traveled equal $t_I > 0$.

The market share of I^a on this Hotelling market is given by (see, for instance, Tirole, 1988):[16]

$$\tfrac{1}{2} + \hat{x}^a = \tfrac{1}{2} + \frac{\int_0^{v^a} (e+v) f(v) dv - \sigma^a - (\int_0^{v^b} (e+v) f(v) dv - \sigma_b)}{2t_I} \tag{43}$$

Hence with Hotelling competition, we find the following expressions: $\hat{x}^a_{V^a} = 1/(2t_I), \hat{x}^a_{V^b} = -1/(2t_I), \hat{x}^a_{V^a V^a} = 0$.

As shown in appendix B, we find the following for the Hotelling equilibrium.

Lemma 6. *In the Hotelling game, we find the following results. For I^a (with similar expressions for I^b) the equilibrium price cost margin is given by*

$$\mu^a = \sigma^a - F(v^a)p^a = t_I + \frac{\int_0^{v^a} (e+v) f(v) dv - F(v^a)p^a - (\int_0^{v^b} (e+v) f(v) dv - F(v_b)p_b)}{3} \tag{44}$$

equilibrium market share

$$\tfrac{1}{2} + x^a = \tfrac{1}{2} + \frac{\int_0^{v^a} (e+v) f(v) dv - F(v^a)p^a - (\int_0^{v^b} (e+v) f(v) dv - F(v_b)p_b)}{6t_I} \tag{45}$$

and equilibrium profits

$$\frac{1}{2t_I}\left(t_I + \frac{\int_0^{v^a} (e+v) f(v) dv - F(v^a)p^a - (\int_0^{v^b} (e+v) f(v) dv - F(v_b)p_b)}{3}\right)^2 \tag{46}$$

Hence, in terms of lemma 2, with Hotelling we have $\psi = 1/3$.[17]

Moving back to before the competitive situation is realized on the health insurance market, I^a's expected market share is given by

$$ms^a = \tfrac{1}{2} + \theta_I x^a = \tfrac{1}{2} + \theta_I \frac{\int_0^{v^a} (e+v) f(v) dv - F(v^a)p^a - (\int_0^{v^b} (e+v) f(v) dv - F(v_b)p_b)}{6t_I} \tag{47}$$

To derive P^1's expected market share, first consider the case where –with probability θ_P– there is Hotelling competition between providers P^1, P^2. Conditional on the patient having chosen insurer I^a, P^1's market share can be derived as follows. As with insurer competition, we assume that $P^{1,2}$ are on the far left and far right of an Hotelling beach with length 1. The probability distribution of the patient's position on this beach is represented by a uniform

[16] As there is Hotelling competition on the provider market as well –with probability θ_P– and the consumer does not yet know his position on the provider-Hotelling-beach, strictly speaking the value offered by I^a is given by

$$\int_0^{v^a} (e+v) f(v) dv - \theta_P \frac{t_P}{4} - \sigma^a \tag{42}$$

As we focus on a symmetric equilibrium, we simplify notation by ignoring the constant $\theta_P \frac{t_P}{4}$.

[17] This follows from equation (A.4) with $d\mu^a/dV^b = -1/3$ in equation (44).

distribution (with density 1). Hence, the position x of the indifferent patient type is given by the standard expression:

$$V^{a1} - t_P x = V^{a2} - t_P(1-x) \tag{48}$$

Solving for x we get P^1's market share:

$$\tfrac{1}{2} + \hat{x}^{a1} = \tfrac{1}{2} + \frac{V^{a1} - V^{a2}}{2t_P} = \tfrac{1}{2} + F(v^a)\frac{e^1 - e^2}{2t_P} \tag{49}$$

With probability $1 - \theta_P$, the consumer randomly chooses one of the providers; choosing P^1 with probability $\frac{1}{2}$. Hence, conditional on the patient choosing I^a who contracted both providers, the expected market share for P^1 is given by

$$ms^{a1} = \theta_P\left(\tfrac{1}{2} + F(v^a)\frac{e^1 - e^2}{2t_P}\right) + (1-\theta_P)\tfrac{1}{2} = \tfrac{1}{2} + \theta_P F(v^a)\frac{e^1 - e^2}{2t_P} \tag{50}$$

By symmetry, we find $ms^{a2} = 1 - ms^{a1}$. Using $\hat{x}^{a1}_{V^{a1}} = 1/(2t_P)$, we write (22) as

$$\left.\frac{de}{dv}\right|_{e^1=e^2, v^a=v^b} = \frac{\theta_P}{4t_P\gamma''(e)}\frac{d(F(v)R(v))}{dv} \tag{51}$$

and in symmetric equilibrium equation (21) becomes

$$\gamma'(e(v)) = \frac{\theta_P}{2t_P}F(v)R(v) \tag{52}$$

Equilibrium is determined by (39):

$$(e(v) + v - c(v))\frac{1 - \frac{\theta_I}{3}}{2} = \frac{F(v)}{f(v)}\left(-\frac{\theta_I}{6}c'(v) + \frac{R(v)\frac{\theta_P}{t_P} - \frac{1}{2}(1-\theta_I)}{\gamma''(e)}\frac{\theta_P}{4t_P}\frac{d(R(v)F(v))}{dv}\right) \tag{53}$$

Figure 2 plots the equations determining equilibrium and first best outcomes –equation (39) and $v + e^*(v) - c(v) = 0$– and $e(v), e^*(v)$ for the Hotelling set-up. For the figure we work with: $F(v) = v$ for $v \in [0,1]$, $c(v) = 3/2v^2$, $\gamma(e) = 7/4e^2$, $t_I = 1$, $t_p = 0.2$ and $\theta_I = 0.4, \theta_P = 0.13$. For these parameter values, the market equilibrium features over-treatment ($v > v^*$) and under-investment ($e < e^*$).

6. Competition and welfare

In this section we derive the effects of provider and insurer competition on welfare. We start with a benchmark result that derives a condition under which first best can be implemented in the market outcome by setting θ_I, θ_P correctly. If θ_I, θ_P are not at first best levels –which is likely in the real world– the market outcome is in one of four possible situations with over/under-treatment and over/under-investment. We characterize when each of these four cases tends to happen.

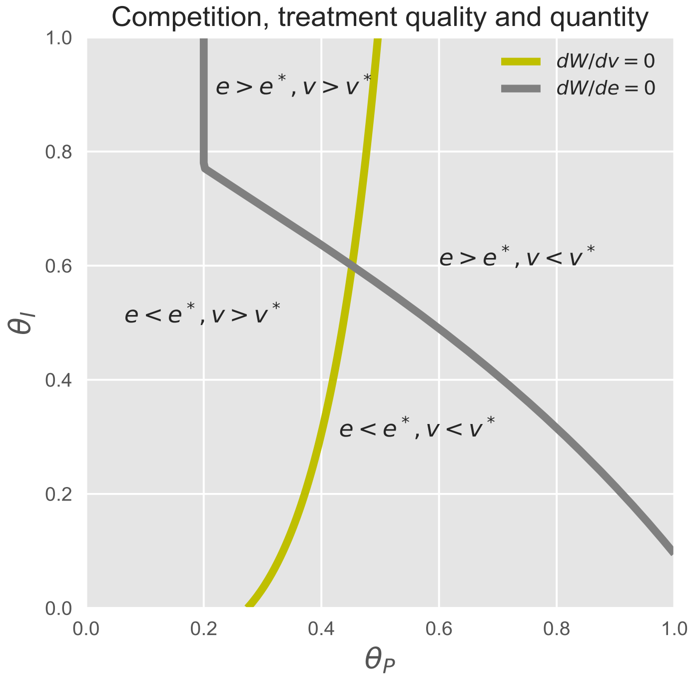

Figure 3: Equilibrium values for dW/de and dW/dv as a function of $\theta_P, \theta_I \in [0, 1]$.

6.1. first best

The following proposition shows that the market can implement first best if the provider market is sufficiently competitive (in terms of $\hat{x}^{a1}_{V^{a1}}$) to generate first best quality and if the insurance market features a strong enough competition effect to offset the private over-investment in quality effect.

Proposition 2. *There exist $\theta_I^*, \theta_P^* \in \langle 0, 1]$ such that the solution to (39) yields v^* and $e^* = e(v^*)$ if and only if in first best with $v = v^*, e = e^*$ it is the case that*

$$\left.\frac{de}{dv^a}\right|_{v=v^*} \leq DC(v^*) \tag{54}$$

and

$$\hat{x}^{a1}_{V^{a1}} R(v^*) \geq \tfrac{1}{2} \tag{55}$$

The intuition for the two conditions is as follows. Starting with (55), if there is little incentive to compete in the provider market ($\hat{x}^{a1}_{V^{a1}} R(v^*)$ is low; say, due to high (perceived) provider switching costs), then even $\theta_P = 1$ is not going to induce first best effort levels. However, if this competition incentive is high enough, we can find $\theta_P^* \leq 1$ such that first best effort can be sustained in the market outcome. Comparing (4) and (21) gives (55).

To understand (54), first note that in first best we have PMVQ$\propto \partial \Pi^a / \partial e < 0$: compare (4) and (28). Because I^a overlooks the positive externality of e on I^b, with first best e^* there is over-investment in quality from I^a's point of view. If effort reacts strongly to v^a –equation

(22)– I^a wants to reduce v^a below v^* to reduce e. The counteracting effect is that insurer competition increases treatment volume: DC(v^*). If the former effect is too big compared to the latter, we cannot find $\theta_I \leq 1$ to implement first best. However, if the competition effect on treatment volume is big enough compared to the effect of v^a on e, first best can be implemented at $\theta_I^* \leq 1$.

In plain words, there should be enough competition incentives for providers to implement first best quality. Further, insurer competition should be intense enough to prevent a reduction in volume in response to the over-investment (from the insurer's point of view) with first best quality. If both conditions are satisfied, we can find θ_I^*, θ_P^* and implement first best volume and quality in the market outcome.

The first point to note is that competition is necessary in both markets. No competition in the provider market ($\theta_P = 0$) leads to under-investment in treatment quality. Providers need an incentive to attract patients in order to invest effort in quality. As the information rent $R(v^*)$ increases, less provider competition is needed to implement first best. In other words, for treatments where the asymmetric information advantage for providers is bigger, the incentive to raise quality is higher (more to be earned from gaining a patient) and provider market competition can be less intense. Further, no competition in the insurance market ($\theta_I = 0$) leads to under-treatment. As mentioned, in first best e^* we have $\partial \Pi^a / \partial e < 0$, hence I^a wants to reduce $v^a < v^*$ in order to reduce e. Insurer competition $\theta_I > 0$ is needed to keep treatment volume at v^*.

In other words, in a setting with non-contractible quality where patients can learn from each others' experiences –such that providers compete for them based on quality– the market can be fine-tuned to outperform a public health care system with no competition between providers. Another implication is that there can be "too much" provider competition. As the provider market becomes more competitive, quality investments increase, but there can be over-investment leading to "gold plating".

As a caveat, the result that the right provider competition intensity can implement first best quality depends on the assumption that patient utility and social welfare are aligned. If this is not the case, provider competition is less likely to be a good policy instrument. Think here of the case where patients value gourmet meals in hospitals while the planner only values improvement in health. Alternatively, the planner would like to contain macro health care costs, say by introducing gatekeepers in the system, while the patient would like to receive any treatment that he thinks can be helpful. Then a socially optimal gatekeeper will lose market share to a physician who is "more generous" from a patient's point of view. With a wedge between patient utility and welfare, competition does not necessarily enhance welfare.

6.2. four cases

If the market outcome is not first best, we are in one of the four cases in table 1. To characterize whether the market outcome features over/under-treatment, we use equation (39) where a positive (negative) right hand side implies under (over) treatment in equilibrium. Further, from (4) with (21) we can write the wedge on e as

$$\frac{dW}{de} = F(v) - 2\gamma'(e) = F(v)(1 - 2\theta_P \hat{x}^{a1}_{V^{a1}} R(v)) \tag{56}$$

Hence, $1 - 2\theta_P \hat{x}^{a1}_{V^{a1}} R(v) > 0$ implies under-investment in quality: an increase in e raises welfare W.

		investment	
		under	over
treatment	under	θ_I low,θ_P medium	θ_I medium,θ_P high
	over	$\theta_I > 0$,θ_P low	θ_I high,θ_P medium

Table 1: Investment and treatment decisions for different intensities of provider and insurer competition.

Table 1 relates over/under investment and over/under treatment to different competition intensities in the insurance and provider markets. Figure 3 is based on the example and parameter values in section 5, where both θ_I and θ_P vary between 0 and 1. The figure shows that each of the four cases (with under/over investment and treatment) can happen.

The figure calculates the equilibrium for each of the combinations θ_P, θ_I. Then it calculates, for this equilibrium, the values of dW/dv and dW/de. The combinations of θ_P, θ_I leading to efficient treatment decisions ($dW/dv = 0$) and efficient quality ($dW/de = 0$) are indicated by the two lines. At the intersection of these two lines, we have a combination of θ_P, θ_I where $dW/dv = dW/de = 0$. These are the first best values θ_P^*, θ_I^* characterized in proposition 2. The four areas bordered by these two lines indicate over/under treatment and investment. We write –with abuse of notation– $v < v^*$ in case $dW/dv > 0$ and $e < e^*$ in case $dW/de > 0$; and similarly for $dW/dv, dW/de < 0$.

under-investment and under-treatment

The first case arises if $\theta_I = 0$ (or, by continuity, θ_I small) and $1 > 2\theta_P \hat{x}^{a1}_{V^{a1}} R(v) > \frac{1}{2}$ (medium range for θ_P): lower right part of figure 3. Because $2\theta_P \hat{x}^{a1}_{V^{a1}} R(v) < 1$, equation (56) implies that there is under-investment. With $\theta_I = 0$ (or small), the competition effect –leading to over-treatment– is dominated by the $\partial \Pi^a / \partial e < 0$ effect. From (21) and (28), we have that

$$\frac{\partial \Pi^a}{\partial e} = \tfrac{1}{2}(1 - \theta_I) F(v) - 2\gamma'(e) < F(v)(\tfrac{1}{2} - 2\theta_P \hat{x}^{a1}_{V^{a1}} R(v)) < 0 \tag{57}$$

To reduce this over-investment (from the insurer's point of view) in effort, insurers reduce treatment v below first best: under-treatment. If, in contrast, it would be the case that $2\theta_P \hat{x}^{a1}_{V^{a1}} R(v) < \frac{1}{2}$, there is under-investment from I^a's point of view, leading to over-treatment to stimulate quality.

Intuitively, with low competition intensity in both markets, there is under-investment in quality. But with $\theta_P \hat{x}^{a1}_{V^{a1}} R(v)$ high enough compared to θ_I, the investment effect exceeds the competition effect and there is under-treatment to reduce over-investment in effort from a private point of view ($\partial \Pi^a / \partial e < 0$).

over-investment and under-treatment

This case arises if $2\theta_P \hat{x}^{a1}_{V^{a1}} R(v) > 1$ causing over-investment from a welfare point of view and hence also from a private point of view. In figure 3 the $dW/dv = 0$ curve is steep above θ_I^*; hence, only for a small range of $\theta_P > \theta_P^*$ there is an upper-bound on θ_I to stay in the region with $v < v^*$. For higher values of θ_P, e is so high and hence $\partial\Pi^a/\partial e$ so negative that there is under-treatment even if $\theta_I = 1$.

Intuitively, with very intense provider competition, providers over-invest in quality to attract patients. With low insurer competition, the investment effect on treatments exceeds the competition effect. This results in under-treatment. But with θ_I too low we get both under-treatment and under-investment.

over-investment and over-treatment

With $\theta_I = 1$ (and, by continuity, $\theta_I < 1$ high) and $\frac{\psi}{2}\frac{c'(v)}{de/dv} > 2\theta_P \hat{x}^{a1}_{V^{a1}} R(v) > 1$, we get both over-investment and over-treatment. The over-investment result follows immediately from $2\theta_P \hat{x}^{a1}_{V^{a1}} R(v) > 1$. To get over-treatment, we need that the competition effect exceeds the (private) over-investment effect. Basically, this implies that $\psi c'(v)$ should be high enough to exceed the over-investment effect. This is the area at the top and in the middle of figure 3.

Intuitively, with intense provider competition there is over-investment in quality to attract patients. Further, with high insurer competition, the competition effect on treatment is high and exceeds the investment effect: over-treatment.

under-investment and over-treatment

Finally, consider the case with $\theta_P = 0$ (by continuity $\theta_P > 0$ small) and $\theta_I > 0$. With low provider competition, providers under-invest in quality. This can happen to the extent that $\partial\Pi^a/\partial e > 0$. In this case, insurers induce higher than efficient treatment v to increase investment (as $de/dv > 0$). And $\theta_I > 0$ adds the competition effect which further raises treatment volume. This is the area on the left of figure 3.

Intuitively, low θ_P leads to under-investment in quality. If profits become increasing in quality, this together with the competition effect leads to over-treatment.

This last case is relevant for the debate whether competition can work at all in health care markets. As mentioned in the Introduction, one concern is that market competition leads to over-treatment (this is how providers earn money) and under-investment in quality (to keep costs low for insurers). However, this last case shows that this outcome is actually caused by lack of competition. More intense provider competition would both raise quality and reduce treatment volume.

7. Diagnosis and remedy

Suppose one concludes that the health care market is not in first best; that is, the market is in one of the four cases above. How should competition intensities θ_I, θ_P change to increase welfare? This section derives two results. First, if the health care market is not in first best, how can we diagnose whether competition intensities are too high or too low? Second, arguably the provider market is especially hard for patients to understand. A patient who does not yet know or understand what ails him, can he decide which hospital will provide the best treatment? Hence, suppose it is simply not feasible to make the provider market competitive enough to get first best effort levels. How should competition in the insurer market be adjusted to compensate for this? We conclude with a discussion of policy instruments to change competition intensities in insurer and provider markets.

7.1. competition up or down?

The following result gives a simple formalization of the idea that under-investment calls for more intense provider competition while under-treatment is remedied with more intense insurer competition.

Proposition 3. *Assume there exist θ_P^*, θ_I^* that implement first best. Then $\theta_P < \theta_P^*$ if and only if*

$$\theta_P R(v^*)\hat{x}^{a1}_{V^{a1}} < \tfrac{1}{2} \tag{58}$$

and $\theta_I > \theta_I^*$ if and only if

$$\theta_I \psi c'(v^*) > (1+\theta_I)\left.\frac{de}{dv}\right|_{v=v^*} \tag{59}$$

where both conditions are evaluated at v^*, e^*.

If first best competition intensities exist, then under-investment in quality (see (56) evaluated at v^*) implies that provider competition is less intense than θ_P^*. Hence, under-investment calls for more intense provider competition (over-investment for less intense provider competition).

Second, if the competition effect –left hand side of (59)– exceeds the private over-investment effect –right hand side– then we have over-treatment and insurer competition should be reduced.

Finally, one may worry that the provider market is so hard to understand for patients that it cannot be competitive enough to implement efficient quality investments. The following result shows that in this case, low provider competition should be compensated by more intense insurer competition.

Proposition 4. *Assume $\hat{x}^{a1}_{V^{a1}} > 0$ close to 0. Then optimal θ_I is set at a higher level than the one that implements efficient treatment volume. The planner induces over-treatment.*

With such low $\hat{x}^{a1}_{V^{a1}}$ we have under-investment in quality even if $\theta_P = 1$. Hence, it is not possible to implement efficient investments. In order to raise investments in the market, the planner increases v beyond its efficient level. This over-treatment is established by increasing insurer competition θ_I beyond θ_I^*. Recall that for $e > 0$ small enough, we have $\partial\Pi^a/\partial e > 0$ and the insurer already induces over-treatment to increase quality: $\theta_I DC + PMVQde/dv^a > 0$ in equation (39). The planner strengthens this effect by raising θ_I beyond θ_I^*. This outcome features over-treatment and under-investment.

7.2. policy instruments

Suppose that from the analysis above, a government concludes that competition intensity should be changed in the provider and/or insurance markets; how can it affect these competition intensities? Here we give examples of policy instruments that can change health care competition intensities.

First, a well known instrument to make patients more sensitive to quality differences between providers, is to publish patient satisfaction reports. By providing patients with more information, they are better able to evaluate provider quality differences thereby increasing $\hat{x}^{a1}_{V^{a1}}$ (Dafny and Lee, 2016).

Another way to increase provider competition is to mandate referring physicians to give patients –at least– five choices for hospitals to get treatment. As reported by Gaynor et al. (2016), this policy was introduced in the UK in 2006 and increased provider competition.

Note that in our paper, (provider) competition is not equated with profits or rents. Sometimes low profits are seen as a signal of intense competition, which is then interpreted as a good thing. But equation (55) implies that in our framework rents can be too low. The reason is that here rents are the result of asymmetric information between the provider and the insurer/planner. By introducing more information, these rents can be reduced.[18] The results above suggest that such improvements in information should go hand-in-hand with more intense provider competition to keep effort at the desired level. Similarly, higher rents can compensate for provider competition that is too slack: as $R(v^*)$ increases, θ^*_P falls (see equation (A.23) in the appendix).

Third, reducing concentration in provider and insurance markets by abolishing mergers also helps to keep these markets competitive. See Haas-Wilson (2003), Gaynor et al. (2015) for discussions of merger waves in US health care markets. Also in the Netherlands, the health insurance and provider markets are highly concentrated due to past mergers (Gaynor et al., 2015, pp. 242).

Related to this, Dafny and Lee (2016) suggest an increase in funding for competition agencies and regulators overlooking health care markets. Such funding also helps authorities to prevent exclusionary and collusive behaviour. Physicians may propose quality standards or certification that keep entrants out of the market. Certain specialties may agree on a “fair” price to charge for their services etc. Such practices can reduce competition intensity and well funded authorities are needed to prevent these.

Finally, three policy instruments related to health insurance competition are the following. First, the government can set the basic health insurance package. This is, for instance, the case in the Netherlands. The set of treatments covered by (private) insurers is then the same. Further, the Dutch government sets the maximum and minimum deductible (difference between these is 500 euro per year). Consequently, the products on the basic insurance market are rather homogeneous and the market is transparent. In other words, insurers cannot reduce competition through product differentiation. See Ericson and Starc (2016) for an analysis of a similar policy on the Massachusetts’ Health Insurance Exchange.

Second, the government can use risk adjustment to prevent insurer competition being distorted by cherry picking of insured. If the differences between customers’ expected profits are small, insurers can focus on competing on a level playing field instead of investing effort in risk selection.

Third, a factor limiting insurer competition is the requirement that an insurer keeps a reserve for each of its insured. If an insurer’s reserves are limited, this constrains the growth in market share that it can accommodate. Reducing the reserve requirement then makes the insurer market more competitive (but also more risky from a financial point of view).

[18] To illustrate, suppose initially a treatment can take values $v \in [0, \bar{v}]$ and optimally patients with $v \in [0, v^a]$ are treated; that is, the insurer pays $p^a = c(v^a)$ per treatment. By investing in monitoring technology the insurer can distinguish treatments $v < v^1$ from treatments $v > v^1$ with $v^1 < v^a$. This implies that only $p^1 = c(v^1) < p^a$ needs to be paid for treatments $v \in [0, v^1]$ which reduces the information rent R.

8. Discussion

We conclude the paper with a discussion of policy implications and extensions of our model.

8.1. policy implications

This paper is motivated by the question whether markets with profit maximizing agents can be used to organize health care. We focus on two contractual imperfections: (i) the cost and value of treatment for a particular patient cannot be contracted and (ii) treatment quality cannot be contracted. We analyze the effects of changes in provider and insurer competition on treatment volume, quality and hence welfare.

We show that if both markets can be made competitive enough, first best volume and quality can be achieved by fine tuning competition intensities in the provider and insurance markets. In both markets (strictly positive) competition is needed to maximize welfare. In the provider market competition is needed to stimulate non-contractable investments in quality. In the insurer market, the fact that investments are not contractable makes quality non-excludable for insurers. Implementing high (first best) quality then leads to over-investment from the insurers' point of view. Insurers would like to reduce the fee-for-service, thereby reducing rents and providers' incentives to invest in quality. To prevent this, insurer competition is necessary to keep volume, prices and provider rents at the right level.

Broadly speaking, under-investment in quality calls for more intense provider competition. Under-treatment can be resolved by increasing insurer competition.

If, however, patients are not rational enough to generate sufficient competition between providers, there will be under-investment and if insurers' profits are increasing in treatment quality there will be over-treatment.

Critics of organizing health via markets sometimes argue that market competition leads to over-treatment (this is how physicians make money) and under-investment (to cut costs). However, we show that this outcome is actually caused by lack of competition; not by an excess of it.

We derived these results in a tractable industrial organization model emphasising the importance of information rents to give providers an incentive to compete in treatment quality. If this information rent is reduced –say, through generating more contractible patient information– provider competition needs to be intensified to keep up the quality incentives.

Equilibrium volume and quality are determined by three effects: treatment efficiency, an insurer competition effect leading to higher volume and the insurers' private valuation of treatment quality. Due to non-excludability of quality investments, insurers value quality less than the social planner.

8.2. extensions

As this is the first model analyzing the interplay between provider and insurer competition, we have kept other dimensions as simple as possible. We finish with a discussion of three natural extensions.

First, we assumed that consumers and firms are rational in the sense of utility and profit maximizing. However, we know from, for instance, Handel and Kolstad (2015) that consumers may not be best described as "fully rational utility maximizing" when choosing insurance. In fact, they may have biases other than not always carefully comparing the different suppliers available, which we focus on here. Also on the provider side there can be other objectives than profit maximization, e.g. providers may be intrinsically motivated to provide quality and they can be not-for-profit. The question then becomes: how does the optimal competition intensity in the insurer and provider markets change taking this into account? E.g. when should competition be relaxed compared to the rational benchmark derived in this paper.

As mentioned in the introduction, we focus on symmetric equilibria where each insurer contracts each provider. This can be generalized to analyze the effects of network formation. The type of question that can be analyzed then: should competition intensities be reduced to avoid networks that become too narrow?

Third, our model does feature moral hazard (on the provider side; possibly leading to over-treatment) but not adverse selection. Introducing two types of patients to see how competition and selection interact would be a generalization. Interesting is the observation by Gaynor et al. (2015, pp. 249) that more seriously ill patients are more sensitive to quality. E.g. chronically ill have more experience of providers and hence know better the quality differences. In this sense, this market segment of the provider market is more competitive. However, high risk insured are not attractive for insurance companies (unless there is a good risk adjustment system). Further, depending on the financing structure used by insurers, these patients may not be profitable for providers either. Does then an increase in competition (on the provider and insurer markets; say by blocking mergers to prevent a fall in competition) still lead to higher quality efforts? Or do providers reduce efforts to avoid patients with high expected costs?

References

Jason Abaluck and Jonathan Gruber. Evolving choice inconsistencies in choice of prescription drug insurance. *American Economic Review*, 106(8):2145–84, August 2016. doi: 10.1257/aer.20130778. URL http://www.aeaweb.org/articles?id=10.1257/aer.20130778.

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

Jan Boone. Competitive pressure: The effects on investments in product and process innovation. *The RAND Journal of Economics*, 31(3):549–569, 2000. ISSN 07416261.

Jan Boone. A new way to measure competition. *The Economic Journal*, 118(531):1245–1261, 2008.

Jeremy I. Bulow, John D. Geanakoplos, and Paul D. Klemperer. Multimarket oligopoly: Strategic substitutes and complements. *Journal of Political Economy*, 93(3):488–511, 1985.

C. Capps, D. Dranove, and M. Satterthwaite. Competition and market power in option demand markets. *RAND Journal of Economics*, 34(4):737–763, Winter 2003.

J. Cohn. *Sick: the untold story of America's health care crisis–and the people who pay the price.* Harper Perennial, 2007.

A. Collard-Wexler, G. Gowrisankaran, and R. Lee. 'Nash-in-Nash' bargaining: A microfoundation for applied work. *Journal of Political Economy*, (forthcoming), 2017.

Leemore Dafny. Estimation and identification of merger effects: An application to hospital mergers. *The Journal of Law and Economics*, 52(3):523–550, 2009.

Leemore S. Dafny and Thomas H. Lee. Health care needs real competition. *Harvard Business Review*, 94(12):76–87, December 2016.

R.A. Deyo and D.L. Patrick. *Hope or hype: the obsession with medical advances and the high cost of false promises*. Amacom, 2005.

M. Drummond, M. Sculpher, G. Torrance, B. O'Brien, and G. Stoddart. *Methods for the economic evaluation of health care programmes*. Oxford University Press, third edition edition, 2005.

Keith M. Marzilli Ericson and Amanda Starc. How product standardization affects choice: Evidence from the massachusetts health insurance exchange. *Journal of Health Economics*, 50:71 – 85, 2016. ISSN 0167-6296.

Joseph Farrell and Paul Klemperer. Chapter 31 coordination and lock-in: Competition with switching costs and network effects. volume 3 of *Handbook of Industrial Organization*, pages 1967 – 2072. Elsevier, 2007.

D. Fudenberg and J. Tirole. *Game theory*. MIT Press, 1991.

Drew Fudenberg and Jean Tirole. The fat-cat effect, the puppy-dog ploy, and the lean and hungry look. *The American Economic Review*, 74(2):361–366, 1984.

Esther Gal-Or. Exclusionary equilibria in health-care markets. *Journal of Economics & Management Strategy*, 6(1):5–43, 1997.

Martin Gaynor, Kate Ho, and Robert J. Town. The industrial organization of health-care markets. *Journal of Economic Literature*, 53(2):235–84, June 2015.

Martin Gaynor, Carol Propper, and Stephan Seiler. Free to choose? reform, choice, and consideration sets in the english national health service. *American Economic Review*, 106 (11):3521–57, November 2016.

M.R. Gold, J.E. Siegel, L.B. Russell, and M.C. Weinstein. *Cost-effectiveness in health and medicine*. Oxford University Press, 1996.

Gautam Gowrisankaran, Aviv Nevo, and Robert Town. Mergers when prices are negotiated: Evidence from the hospital industry. *American Economic Review*, 105(1):172–203, 2015.

D. Haas-Wilson. *Managed care and monopoly power: the antitrust challenge*. Harvard University Press, 2003.

Benjamin R. Handel and Jonathan T. Kolstad. Health insurance for "humans": Information frictions, plan choice, and consumer welfare. *American Economic Review*, 105(8):2449–2500, 2015.

K. Ho. Insurer-provider networks in the medical care market. *American Economic Review*, 99 (1):393–430, 2009.

Kate Ho and Robin S. Lee. Insurer competition in health care markets. *Econometrica*, 85(2): 379–417, 2017. ISSN 1468-0262.

Katherine Ho. The welfare effects of restricted hospital choice in the us medical care market. *Journal of Applied Econometrics*, 21(7):1039–1079, 2006.

Jonathan T. Kolstad and Michael E. Chernew. Quality and consumer decision making in the market for health insurance and health care services. *Medical Care Research and Review*, 66 (1_suppl):28S–52S, 2009. PMID: 19029288.

J.E. McDonough. *Inside national health reform*. University of California Press, 2011.

Thomas G. McGuire. Chapter five - demand for health insurance. In Thomas G. Mcguire Mark V. Pauly and Pedro P. Barros, editors, *Handbook of Health Economics*, volume 2 of *Handbook of Health Economics*, pages 317 – 396. Elsevier, 2011.

John A. Nyman. The value of health insurance: the access motive. *Journal of Health Economics*, 18(2):141 – 152, 1999.

OECD. *Fiscal Sustainability of Health Systems*. OECD Publishing, 2015.

D. G. Pope. Reacting to rankings: Evidence from "america's best hospitals". *Journal of Health Economics*, 28(6).:1154–1165., 2009.

Michael E. Porter and Elizabeth Olmsted Teisberg. *Redefining health care: creating value-based competition on results*. Harvard Business School Press, Boston, Massachusetts, 2006.

Carol Propper, Simon Burgess, and Denise Gossage. Competition and quality: Evidence from the nhs internal market 1991-9. *The Economic Journal*, 118(525):138–170, 2007.

C Schoen, R Osborn, D Squires, M M Doty, R Pierson, and S Applebaum. How Health Insurance Design Affects Access To Care And Costs, By Income, In Eleven Countries. *Health Affairs*, 29(12):1–12, 2010.

Cathy Schoen, Sara R Collins, Jennifer L Kriss, and Michelle M Doty. How many are underinsured? trends among u.s. adults, 2003 and 2007. *Health affairs (Project Hope)*, 27(4): 298–309, 2008.

Erik Schokkaert and Carine van de Voorde. Chapter 15 - user charges. In S. Glied and P. Smith, editors, *Oxford Handbook of Health Economics*, pages 329 – 353. Oxford University Press, 2011.

I. Segal. Contracting with externalities. *Quarterly Journal of Economics*, 114:337–388, 1999.

Sophie Terp, Brandon Wang, Brian Raffetto, Seth A. Seabury, and Michael Menchine. Individual physician penalties resulting from violation of emergency medical treatment and labor act: a review of office of the inspector general patient dumping settlements, 2002-2015. *Academic Emergency Medicine*, 24(4):442–446, 2017.

Jean Tirole. *The theory of industrial organization*. MIT Press, 1988.

W. P.M.M. van de Ven and F. T. Schut. Universal mandatory health insurance in the netherlands: A model for the united states? *Health Affairs*, 27(3):771–781, 2008.

Wynand P.M.M. van de Ven and Frederik T. Schut. Chapter 17 - guaranteed access to affordable coverage in individual health insurance markets. In S. Glied and P. Smith, editors, *Oxford Handbook of Health Economics*, pages 380 – 404. Oxford University Press, 2011.

J.E. Wennberg. *Tracking Medicine*. Oxford University Press, 2010.

A. Proof of results

Proof of lemma 2 The first order condition for μ^a can be written as

$$\tfrac{1}{2} + \hat{x}^a - \mu^a \hat{x}^a_{V^a} = 0 \tag{A.1}$$

which we can solve for μ^a as

$$\mu^a = \frac{\frac{1}{2} + \hat{x}^a}{\hat{x}^a_{V^a}} \tag{A.2}$$

Using an envelope argument and equation (A.2), we can write

$$\pi^{ca}_{V^a} = \mu^a \left(\hat{x}^a_{V^a} - \hat{x}^a_{V^b} \frac{d\mu^b}{dV^a} \right) = (\tfrac{1}{2} + x^a)(1 - \psi) \tag{A.3}$$

with

$$\psi = \frac{\hat{x}^a_{V^b}}{\hat{x}^a_{V^a}} \frac{d\mu^b}{dV^a} \tag{A.4}$$

From $\hat{x}^a + \hat{x}^b = 0$ and symmetry we find that $\hat{x}^a_{V^b} = -\hat{x}^a_{V^a}$ and hence $\psi = -d\mu^b/dV^a$.

Differentiating the first order condition (A.1) with respect to V^b gives

$$\frac{d\mu^a}{dV^b} = \frac{\hat{x}^a_{V^b} - \mu^a \hat{x}^a_{V^a V^b}}{2\hat{x}^a_{V^a} - \mu^a \hat{x}^a_{V^a V^a}} \in \langle -1, 0 \rangle \tag{A.5}$$

The denominator is positive as $\hat{x}^a_{V^a} > 0, \hat{x}^a_{V^a V^a} \leq 0$ by assumption. The fact that the denominator is positive also implies that the second order condition for μ^a is indeed satisfied. The numerator is negative as $\hat{x}^a_{V^b} < 0$ by assumption and $\hat{x}^a_{V^a V^b} \geq 0$. This last inequality follows from differentiating $\hat{x}^a_{V^b} = -\hat{x}^a_{V^a}$ with respect to V^a and $\hat{x}^a_{V^a V^a} \leq 0$. Hence, we find that $d\mu^a/dV^b < 0$. Finally, $d\mu^a/dV^b > -1$ can be written as

$$0 = \hat{x}^a_{V^a} + \hat{x}^a_{V^b} > \mu^a(\hat{x}^a_{V^a V^a} + \hat{x}^a_{V^a V^b}) - \hat{x}^a_{V^a} = -\hat{x}^a_{V^a} \tag{A.6}$$

which holds as $\hat{x}^a_{V^a} > 0$.Hence, we find that $\psi = -\frac{d\mu^b}{dV^a} \in \langle 0, 1 \rangle$.

Since we assume that $\hat{x}^a$ is continuously differentiable, it is continuous in V^a, V^b. As $\hat{x}^a(V^a, V^b) = -\frac{1}{2}$ for $V^a < 0$, we can limit ourselves to $\mu^a \in [0, V^a]$ which is a compact set. Profits are concave in μ^a (as $\pi^a_{\mu^a \mu^a} = -2\hat{x}^a_{V^a} + \mu^a \hat{x}^a_{V^a V^a} \leq 0$) and hence a pure strategy Nash equilibrium exists (see e.g. Fudenberg and Tirole, 1991, Theorem 1.2). □

Proof of lemma 3 By the assumptions on $x^a, \hat{x}^{a1}$ etc. and γ, the function $\hat{\pi}^1$ in (17) is continuous in e^1. Further, as $\lim_{e \to +\infty} \gamma'(e) = +\infty$ while $x^a, \hat{x}^{a1}$ etc lie in $[-\frac{1}{2}, \frac{1}{2}]$, there exists $\bar{e}$ such that P^1 never chooses $e^1 > \bar{e}$. Hence, without loss of generality we can assume that e^1 is chosen from a compact set $[0, \bar{e}]$. By our assumption that $\hat{\pi}^1$ is concave in e^1 off the equilibrium path (and because of (A.10) below $\hat{\pi}^1$ is concave in e^1 on the equilibrium path), an equilibrium in pure strategies exists (Fudenberg and Tirole, 1991, Theorem 1.2). We focus on a symmetric equilibrium given by equation (21). By our assumptions on $\hat{x}^{a1}$ and γ, this equation has a solution for every value of v. This can be seen as follows. If $\gamma'(0) = 0$ equals the right hand side of (21) evaluated at 0, then the (corner) solution is $e = 0$. If $\gamma'(0)$ is smaller than this right hand side, then there exists an $e > 0$ such that this equation holds. This follows from the following three observations: (i) $\gamma''(e) > 0$, (ii) $\lim_{e\to+\infty}\gamma'(e) = +\infty$ and (iii) the right hand side of this equation does not vary with e. This last point can be seen as follows. We know that

$$\hat{x}^{a1}_{V^{a1}} + \hat{x}^{a1}_{V^{a2}} = \hat{x}^{a1}_{V^{a1}} + \hat{x}^{a2}_{V^{a1}} = 0 \tag{A.7}$$

where the first equality follows from symmetry and the second equality follows from our assumption that the size of the market is fixed (patient always visits a provider). As this equation holds for all V^{a1}, we can differentiate with respect to V^{a1} to find

$$F(v^a)(\hat{x}^{a1}_{V^{a1}V^{a1}} + \hat{x}^{a1}_{V^{a1}V^{a2}}) = 0 \tag{A.8}$$

which implies that the right hand side of (21) does not vary with e.

Let SOC denote the second order condition for e^1, then we can take the derivative of (20) which we evaluate at the symmetric equilibrium to arrive at:

$$\begin{aligned}(-\mathrm{SOC})\frac{de^1}{dv^a} &= R^a_{v^a}\tfrac{1}{2}\theta_P\hat{x}^{a1}_{V^{a1}}F(v^a)\\ &+ R^a[\tfrac{1}{2}\theta_I f(v^a)x^a_{V^a}F(v^a) + \theta_I x^a_{v^a}\theta_P\hat{x}^{a1}_{V^{a1}}F(v^a) + \tfrac{1}{2}\theta_P\hat{x}^{a1}_{V^{a1}v^a}F(v^a) + \tfrac{1}{2}\theta_P\hat{x}^{a1}_{V^{a1}}f(v^a)]\\ &+ R^b[-\tfrac{1}{2}\theta_I f(v^a)x^a_{V^a} - \theta_I x^a_{v^a}\theta_P\hat{x}^{b1}_{V^{b1}}F(v^b)]\\ &= \tfrac{1}{2}\theta_P\hat{x}^{a1}_{V^{a1}}\frac{d(R(v)F(v))}{dv}\end{aligned} \tag{A.9}$$

where we have used that in symmetric equilibrium: $R^a = R(v^a) = R(v^b) = R^b$, $x^a_{V^a} = -x^a_{V^b}$, $x^a_{V^aV^a} = -x^a_{V^aV^b}$, $\hat{x}^{a1}_{v^a} = (\hat{x}^{a1}_{V^{a1}} + x^{a1}_{V^{a2}})(e+v^a)f(v^a) \equiv 0$. Next we derive that

$$\begin{aligned}\mathrm{SOC} &= -\gamma''(e) + R^a[\tfrac{1}{2}\theta_I F(v^a)(x^a_{V^ae^1} + x^a_{V^be^1}) + \theta_I x^a_{e^1}\theta_P\hat{x}^{a1}_{V^{a1}}F(v^a) + \tfrac{1}{2}\theta_P\hat{x}^{a1}_{V^{a1}e^1}F(v^a)]\\ &+ R^b[-\tfrac{1}{2}\theta_I F(v^a)(x^a_{V^ae^1} + x^a_{V^be^1}) - \theta_I x^a_{e^1}\theta_P\hat{x}^{b1}_{V^{b1}}F(v^b) + \tfrac{1}{2}\theta_P\hat{x}^{b1}_{V^{b1}e^1}F(v^b)]\\ &= -\gamma''(e) + \theta_P\hat{x}^{a1}_{V^{a1}V^{a1}}R(v)F(v)^2 < 0\end{aligned} \tag{A.10}$$

where we use that $\hat{x}^{a1}_{V^{a1}e^1} = \hat{x}^{a1}_{V^{a1}V^{a1}}F(v^a) \leq 0$ and in symmetric equilibrium ($v^a = v^b, e^1 = e^2$) a number of terms cancel. Note that SOC < 0 (concavity) is also needed for the first order condition to maximize profits. Combining equations (A.9) and (A.10), we find the expression in (22). □

Proof of lemma 4 In order to derive the capitation payments $T^{a,b}$, we need to specify what happens if an insurer does not contract both providers. In particular, what happens if P^1 rejects I^a's contract. We resolve this out-of-equilibrium reaction as follows. As the consumer first chooses an insurer and after this a provider, we assume that insurer preference is stronger than provider preference. If, with probability $\frac{1}{2}(1-\theta^I)$ the consumer prefers I^a, he chooses I^a even if P^1 is not contracted (out of equilibrium). That is, the cost of not choosing I^a in this case is higher than the option value of visiting P^1 by choosing I^b.[19] With probability $1-\frac{1}{2}(1-\theta_I)$ the consumer does not have an explicit preference for I^a. In this case he chooses I^b as there is a probability $\frac{1}{2}(1-\theta_P)$ that he wants to visit P^1 (only).

Given this assumption, we continue by combining (23) and (25) to write

$$\Pi^a = \pi^a + 2\pi^1 - 2\pi^{1,-a} \tag{A.11}$$

[19] If I^a has not contracted any provider, the consumer goes to I^b.

We argue below that $\pi^{1,-a}$ does not depend on v^a; neither directly nor via e^2. Hence, $2\pi^{1,-a}$ is the constant referred to in the lemma. Using (14) and (17), leads to the expression in (26).

The profit of P^1 after rejecting I^a's contract, does not directly depend on v^a as it does not receive a patient with I^a as insurer. P^1 chooses e^1 to maximize

$$\Pi^1 = (1 - \tfrac{1}{2}(1-\theta_I))R^b\left(\tfrac{1}{2} + \theta_P \hat{x}^{b1}(V^{b1}, V^{b2})\right) - \gamma(e^1) - T^b \tag{A.12}$$

which does not depend on v^a if e^2 does not depend on v^a. P^2 chooses e^2 to maximize

$$\Pi^2 = \tfrac{1}{2}(1-\theta_I)R^a + (1 - \tfrac{1}{2}(1-\theta_I))R^b\left(\tfrac{1}{2} + \theta_P \hat{x}^{b2}(V^{b2}, V^{b1})\right) - \gamma(e^2) - T^a - T^b \tag{A.13}$$

Hence, e^2 does not depend on v^a either. Intuitively, when P^1 and P^2 compete in efforts, it is on the I^b market only which is not affected by v^a. Consequently, $\pi^{1,-a}$ does not depend on v^a.

The derivative $\partial\Pi^a/\partial v^a$ can be written as follows, where we evaluate the derivative at $e^1 = e^2 = e$ and $v^a = v^b = v$:

$$\frac{\partial \Pi^a}{\partial v^a} = \tfrac{1}{2}(1-\theta_I)(e+v-c(v))f(v) + \theta_I \pi^{ca}_{V^a}(e+v-c(v))f(v) + F(v)c'(v)(-\tfrac{1}{2}(1-\theta_I) - \theta_I\pi^{ca}_{V^a} + \tfrac{1}{2}) \tag{A.14}$$

which can be written as (27) using the expression for $\pi^{ca}_{V^a}$ in (A.3) with $x^a = 0$ in symmetric equilibrium.

The derivative $\partial\Pi^a/\partial e$ evaluated in symmetric equilibrium can be written as

$$\frac{\partial \Pi^a}{\partial e} = \tfrac{1}{2}(1-\theta_I)F(v) + \theta_I F(v)(\pi^{ca}_{V^a} + \pi^{ca}_{V^b}) - 2\gamma'(e) \tag{A.15}$$

which can be written as (28) since $\pi^{ca}_{V^a} + \pi^{ca}_{V^b} = 0$. This last equality is derived as follows. Following the argument in (A.3), we find that

$$\pi^{ca}_{V^a} = \mu^a\left(\hat{x}^a_{V^a} - \hat{x}^a_{V^b}\frac{d\mu^b}{dV^a}\right) \tag{A.16}$$

$$\pi^{ca}_{V^b} = \mu^a\left(\hat{x}^a_{V^b} - \hat{x}^a_{V^b}\frac{d\mu^b}{dV^b}\right) \tag{A.17}$$

Adding these equations gives

$$\pi^{ca}_{V^a} + \pi^{ca}_{V^b} = -\mu^a \hat{x}^a_{V^b}\left(\frac{d\mu^b}{dV^a} + \frac{d\mu^b}{dV^b}\right) \tag{A.18}$$

because $\hat{x}^a + \hat{x}^b = 0$ and symmetry in the functions x^a, x^b imply $\hat{x}^a_{V^a} + \hat{x}^a_{V^b} = 0$. Equation (A.5) gives $d\mu^a/dV^b$; a similar derivation gives

$$\frac{d\mu^a}{dV^a} = \frac{\hat{x}^a_{V^a} - \mu^a \hat{x}^a_{V^aV^a}}{2\hat{x}^a_{V^a} - \mu^a \hat{x}^a_{V^aV^a}} \tag{A.19}$$

Again using $\hat{x}^a + \hat{x}^b = 0$ and symmetry in the functions x^a, x^b, we find $d\mu^b/dV^a + d\mu^b/dV^b = 0$ and therefore $\pi^{ca}_{V^a} + \pi^{ca}_{V^b} = 0$ as claimed above. □

Proof of proposition 1 Starting from equation (29), we write

$$(e+v-c(v))f(v)\frac{1-\psi\theta_I}{2} + \theta_I\frac{\psi}{2}F(v)c'(v) + (\tfrac{1}{2}(1-\theta_I)F(v) - 2\gamma'(e))\frac{de}{dv^a} = 0 \tag{A.20}$$

Substituting $\gamma'(e)$ from (21) and using the definitions of TE, DC and PMVQ, we find the expression in the proposition.

Note that TE(0)= 0 and $\text{TE}'(v) = e'(v) + 1 - c'(v)$. Then $\text{TE}'(0) = 1$ because $c'(0) = 0$ and $e'(0) = 0$. This last equality follows from (37)

$$\gamma''(e)\frac{de}{dv} = \theta_P \frac{d(R(v)F(v)\hat{x}^{a1}_{V^{a1}})}{dv} = 0 \tag{A.21}$$

at $v = 0$ because $R(0) = F(0) = 0$. Further, DC(0)=0 and $\text{DC}'(v) = \frac{\psi}{2}c''(v) \geq 0$. Finally, $\text{PMVQ}(0) = \frac{1}{2}(1 - \theta_I) \geq 0$ and $de/dv^a|_{v^a=0}= 0$ from (22) because $R(0) = F(0) = 0$.

Hence, starting at $v = 0$, the left and right hand side of (39) are equal to 0. As v increases, the left hand side becomes strictly positive as $\text{TE}'(0) > 0$. The right hand side becomes negative since $\text{DC}(v) \geq 0$ and $\text{PMVQ}de/dv^a \geq 0$ for $v > 0$ close to 0. Consequently, if there is an interior solution to this equation (see lemma 5), the right hand side of (39) cuts the left hand side from below. □

Proof of lemma 5 Similar to the proofs of lemma's 2 and 3, we assume that the insurers' profits are concave in their strategic variables and that without loss of generality these strategic variables can be limited to compact domains. Then an equilibrium in pure strategies exists (Fudenberg and Tirole, 1991, Theorem 1.2).

We know from the proof of proposition 1 that the right hand side of (39) cuts the left hand side from below. There are two possibilities:(i) the left and right hand sides intersect at $v \in \langle 0, \bar{v} \rangle$ and we have an interior solution for v in equilibrium; (ii) the left and right hand side do not intersect. That is, the left hand side lies above the right hand side for all $v \in \langle 0, \bar{v}]$. In this case, there is a corner solution for v; we rewrite equation (39) –evaluated at $v = \bar{v}$– to take this into account:

$$(e(\bar{v}) + \bar{v} - c(\bar{v}))\frac{1 - \psi\theta_I}{2} \geq \frac{F(\bar{v})}{f(\bar{v})}\left(-\theta_I \frac{\psi}{2}c'(\bar{v}) + (2\theta_P R(\bar{v})\hat{x}^{a1}_{V^{a1}} - \tfrac{1}{2}(1 - \theta_I))\left.\frac{de}{dv}\right|_{v=\bar{v}}\right) \tag{A.22}$$

In this case, effort is determined by equation (21) evaluated at $v = \bar{v}$. □

Proof of proposition 2 We want to implement v^*, e^* in the market outcome. Comparing equations (4) and (21), we see that it must be the case that

$$\theta_P \hat{x}^{a1}_{V^{a1}} R(v^*) = \tfrac{1}{2} \tag{A.23}$$

As $\theta_P \in [0, 1]$, this is feasible if (55) holds.

Substituting first best v^*, e^* into (39) yields (A.23) and $TE(v^*) = 0$ and hence

$$-\theta_I DC(v^*) + \tfrac{1}{2}(1 + \theta_I)\left.\frac{de}{dv^a}\right|_{v=v^*} = 0 \tag{A.24}$$

At $\theta_I = 0$, we find that the expression is positive because $\frac{1}{2}de/dv^a > 0$. Hence, we have a value of $\theta_I \leq 1$ where (A.24) holds if and only if at $\theta_I = 1$ it is the case that

$$-DC(v^*) + \left.\frac{de}{dv^a}\right|_{v=v^*} \leq 0 \tag{A.25}$$

This is condition (54). □

Proof of proposition 3 If (58) holds, then $\theta_P < \theta_P^*$. To restore first best e^*, θ_P needs to be increased. Similarly, if (58) does not hold, we have $\theta_P \geq \theta_P^*$.

Evaluating (37) at v^*, e^*, implies equation (A.23) and hence PMVQ$= -\frac{1}{2}(1+\theta_I)$. Then the right hand side of (39) evaluated at v^* is negative if and only if

$$-\theta_I \psi c'(v^*) + (1+\theta_I) \left.\frac{de}{dv}\right|_{v=v^*} < 0 \tag{A.26}$$

which is equivalent to (59). Assume first that (A.26) is negative and thus θ_I needs to be adjusted to increase it. The derivative of (A.26) with respect to θ_I can be written as

$$-\psi c'(v^*) + \frac{1}{2} \frac{\frac{1}{R^*} \left.\frac{d(R(v)F(v))}{dv}\right|_{v=v^*}}{2\gamma''(e^*) - \frac{\hat{x}^{a1}_{V^{a1}V^{a1}}}{\hat{x}^{a1}_{V^{a1}}}} < 0 \tag{A.27}$$

as is implied by (54) –we assume that θ_I^*, θ_P^* exist to implement first best. Hence, we need to reduce θ_I to implement first best; or equivalently $\theta_I > \theta_I^*$. With a similar reasoning, if the expression in (A.26) is positive, we have $\theta_I < \theta_I^*$. □

Proof of proposition 4 For this proof we write the functions determining market shares on the health provider market as $\varepsilon x^{a1}, \varepsilon x^{a2}$ etc. with $\varepsilon > 0$ close to 0. We write

$$\frac{dW}{d\theta_I} = \left(\frac{\partial W}{\partial v} + \frac{\partial W}{\partial e}\frac{de}{dv}\right)\frac{dv}{d\theta_I} \tag{A.28}$$

where $de/dv > 0$ and

$$\frac{\partial W}{\partial e} = F(v)(1 - 2\theta_P R(v)\varepsilon \hat{x}^{a1}_{V^{a1}}) > 0 \tag{A.29}$$

for $\varepsilon > 0$ close to 0. Next, we determine $dv/d\theta_I$. Equilibrium in the market is determined by equation (39) which we write here as follows, using equation (40):

$$(\gamma')^{-1}(\theta_P R(v)F(v)\hat{x}^{a1}_{V^{a1}}) + v - c(v) = \frac{F(v)}{f(v)} \frac{-\theta_I \psi c'(v) + (4\theta_P R(v)\hat{x}^{a1}_{V^{a1}} - (1-\theta_I))\frac{de}{dv}}{1 - \psi\theta_I} \tag{A.30}$$

The left hand side of this equation does not depend on θ_I, hence we only need to derive how the right hand side varies with θ_I:

$$\frac{d\left(\frac{-\theta_I\psi c'(v) + (4\theta_P R(v)\hat{x}^{a1}_{V^{a1}} - (1-\theta_I))\frac{de}{dv}}{1-\psi\theta_I}\right)}{d\theta_I} \approx \frac{-\psi c'(v)}{(1-\psi\theta_I)^2} < 0 \tag{A.31}$$

for $\varepsilon > 0$ close to 0. In figure 2, an increase in θ_I pulls down the right hand side of equation (39) and hence leads to higher v and higher e.

Using these results in equation (A.28), we find at $\theta_P = 1$ and efficient treatment, $e(v) + v - c(v) = 0$, that $dW/d\theta_I > 0$. Hence, the planner increases θ_I beyond the level that induces efficient treatment leading to over-treatment in case $\varepsilon \hat{x}^{a1} > 0$ is close to 0. □

B. Hotelling example

Proof of lemma 6

I^a sets σ^a to solve

$$\max_{\sigma^a} \left(\frac{1}{2} + \frac{\int_0^{v^a} (e+v)f(v)dv - \sigma^a - (\int_0^{v^b} (e+v)f(v)dv - \sigma^b)}{2t_I} \right) (\sigma^a - F(v^a)p^a) \tag{B.32}$$

The first term gives I^a's market share and $\sigma^a - F(v^a)p^a$ is the expected margin μ^a it makes on a sold insurance contract. If the consumer buys I^a's contract, I^a receives the premium σ^a; with probability $F(v^a)$ the consumer is treated by a provider and I^a needs to pay the provider the fee-for-service $p^a = c(v^a)$.

The first order conditions for σ^a, σ^b can be written in matrix notation as:

$$\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} \sigma^a \\ \sigma^b \end{pmatrix} = \begin{pmatrix} t_I + \int_0^{v^a} (e+v)f(v)dv - \int_0^{v^b} (e+v)f(v)dv + F(v^a)p^a \\ t_I + \int_0^{v^b} (e+v)f(v)dv - \int_0^{v^a} (e+v)f(v)dv + F(v^b)p^b \end{pmatrix} \tag{B.33}$$

Inverting the matrix on the left hand side, gives the following expression for σ^a:

$$\sigma^a = t_I + \frac{\int_0^{v^a} (e+v)f(v)dv - \int_0^{v^b} (e+v)f(v)dv + 2F(v^a)p^a + F(v^b)p^b}{3} \tag{B.34}$$

With this it is straightforward to derive the expressions in the lemma. □