Journal of Health Economics 67 (2019) 102222

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

Health provider networks with private contracts: Is there under-treatment in narrow networks?☆

Jan Boone

CentER, TILEC, CEPR, Department of Economics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

ARTICLE INFO

Article history:
Received 14 January 2019
Received in revised form 29 June 2019
Accepted 8 July 2019
Available online 18 August 2019

JEL classification:
I13
I11

Keywords:
Private contracts
Two-part tariffs
Fee-for-service
Capitation
Any Willing Provider laws
Price transparency

ABSTRACT

Contracts between health insurers and providers are private. By modelling this explicitly, we find the following. Insurers with bigger provider networks, pay providers higher fee-for-service rates. This makes it more likely that a patient is treated and hence health care costs and utilization increase with provider network size. Although providers are homogeneous, the welfare maximizing provider network can consist of two or more providers. Provider profits are positive whereas they would be zero with public contracts. Increasing transparency of provider prices increases welfare only if consumers can "mentally process" the prices of all treatments involved in an insurance contract. If not, it tends to reduce welfare.

1. Introduction

There is a general presumption that insurers can keep health care expenditure and costs down by contracting with a small number of providers. This is based on three types of evidence. First, cross section studies like Coe et al. (2013) which compare insurance plans with narrow and broad networks. The latter tend to have higher cost per capita. This type of evidence can be plagued by adverse selection problems (broader networks tend to be more attractive to people with higher expected health care costs). This is not the case with evidence based on country wide changes like managed care in the US. As documented by Cutler (2004), Dranove et al. (1993) and Dranove (2000) the shift from indemnity insurance to managed care where insurers contract a small provider network has reduced the growth in health care expenditure in the US. The backlash which forced insurers to contract broader networks has led to increased health care costs (Lesser et al., 2003).[1] Finally, Klick and Wright (2014) analyze the effect of Any Willing Provider (AWP) laws on health care costs. They show that AWP laws, which make it harder for insurers to contract a narrow provider network, tend to raise health care costs.

It is important to distinguish the effect of network size on health care expenditure on the one hand and on health care utilization and costs on the other. Cutler et al. (2000), analyzing two forms of heart disease, mainly find price effects; i.e. selective contracting and managed care lead to lower expenditure. Zwanziger and Melnick (1988), Zwanziger et al. (1994), Zwanziger and Melnick (1996) and Chernew et al. (2008) document utilization and cost effects. Chernew and Newhouse (2011) give an overview of the effects of managed care on both health care expenditure and costs. In this paper, we focus on the effect of selective contracting on utilization and costs.

We analyse the effect of network size on health care costs and welfare using a framework with private contracting between insur-

☆ Comments and suggestions from two anonymous referees, the Editor Luigi Siciliani, David Bardey and seminar participants in Copenhagen, Siena and at the 16th annual European Health Economics Workshop in Toulouse are much appreciated. Financial support from the Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged.

E-mail address: j.boone@uvt.nl

[1] Two caveats here. First, one should be careful interpreting shifts in the US health care system as being (only) affected by insurers'/HMOs' network size. Second, in terms of terminology, we interpret indemnity insurance as using a network consisting of all providers.

https://doi.org/10.1016/j.jhealeco.2019.102222

ers and providers. Private contracting means that only the parties directly involved in the contracting can observe the contract, outsiders cannot (see, for example, Hart and Tirole, 1990; Segal, 1999).[2] In our context this implies that only the insurer and provider involved know the details of the contract; neither other insurers, providers nor consumers know what is in the contract.

Two reasons to motivate an analysis with private contracting are the following. First, the main reason is that contracts between insurers and providers are, in fact, private due to confidentiality clauses (Muir et al., 2013). Second, public contracts cannot easily explain why a narrow network leads to lower costs. We will review some well known arguments based on public contracting in Section 3. We argue that with public contracts, if there are effects of network size, these mainly involve shifting rents between insurers and providers. Treatment decisions and welfare are unaffected. As a consequence, the models in Section 3 cannot address worries that narrow networks lead to under-treatment (see, for instance, Terhune, 2013; Pear, 2014).

We introduce a model with homogeneous providers treating patients at constant marginal costs. Insurers offer providers contracts privately. These contracts specify two-part tariffs: fee-for-service (variable part) and capitation fee (fixed part). On the health insurance market, insurers offer contracts specifying a co-payment for insured who need treatment. Both the fee-for-service and the co-payment can be used to reduce health care consumption: fee-for-service below treatment cost – treatment itself is not profitable for providers – is known as supply side cost sharing and a positive co-payment as demand side cost sharing (Ellis and McGuire, 1993). In the former case, the capitation fee compensates providers up front for loss-making treatments. The health economics literature tends to focus on the extreme contracts – either pure capitation, no fee-for-service, or pure fee-for-service, no capitation. We allow both instruments to be used simultaneously.[3] The question is: how do provider choice and private contracting affect optimal supply and demand side cost sharing? To illustrate, how can an insurer use supply side cost sharing (like a capitation fee) when patients are free to choose their provider? The main insight is that with private contracting, supply side cost sharing is less attractive for insurers with a broad network.

To make the question more concrete, consider the case of a GP who is partly paid through capitation. In most countries, people register with a GP practice. Hence, a GP knows the potential number of patients that can visit during the year. With competing hospitals this is different. Patients are not required to register ex ante with a provider or specialist for all possible treatments that can be relevant. Thus the potential patient population is not known for a provider. Moreover, strategic choices made by the insurer influence the number of patients that will visit the hospital for treatment. With supply side cost sharing, treatment of these patients is loss making. How can a provider assess whether the capitation fee is high enough to cover expected losses? The strategic uncertainty associated with the number of patients that will visit a provider is the driving force of the model.

In particular, we show the following three results with private contracts and competing providers. First, as the size of the network increases, the fee-for-service goes up as “aggressive” capitation contracts (low fee-for-service, high capitation) become too expensive for insurers. As providers cannot see each others’ contracts, they think that many patients will visit them and therefore providers demand a high capitation.[4] This high capitation fee overcompensates as in equilibrium fewer patients visit the provider. To reduce the high capitation, the insurer raises the fee-for-service (less supply side cost sharing; lower loss per treatment) and a bigger network leads to higher health care utilization and costs. Similarly, AWP laws by expanding the network raise costs as well.

Second, although providers are homogeneous in our model, with private contracts they make positive profits (unless there is either only one provider in the network or the fee-for-service equals the treatment cost). As providers gain from private contracting, the model motivates the use of confidentiality clauses that we see in the real world.[5]

Third, the fee-for-service affects the probability that a patient is treated by a provider (Ellis and McGuire, 1993). Hence, the fee is payoff relevant for a consumer buying insurance. As contracts are private, consumers cannot observe the fee directly. However, network size signals the level of the fee-for-service: a bigger network signals a higher probability of being treated. As we focus on the case with symmetric consumers, we do not need to invoke adverse selection to get this effect of network size on utilization. The equilibrium network size is determined by the trade off between consumer utility and provider profits.

The relevance of these results is the following. Health care costs are rising in almost all OECD countries and governments are eager to reduce these costs. If reducing insurers’ network size helps to contain costs, it may be a promising option. However, as mentioned, people worry that smaller networks tend to cause under-treatment: people that need treatment go without. This paper offers a framework where network size affects health care utilization and costs; thus, we can evaluate the welfare effects of changes in the network.

This paper is related to the following strands of literature. First, the literature on demand and supply side instruments to curb moral hazard (see, for instance Ellis and McGuire, 1993, for an overview). Papers in this literature work with public contracts which has two implications. First, demand and supply side cost sharing can be analyzed separately. Second, the first best outcome is implementable[6] (see Eq. (11) below). Neither implication holds in a model with private contracting.

Second, papers on provider networks and how they are organized include Ma and McGuire (1997, 2002); these papers feature non-contractible physician effort and insurers imposing targets for providers’ supply of care. They characterize the outcomes that can be achieved with optimal contracts. Outcomes are typically not efficient. Papers that allow for provider heterogeneity include Capps et al. (2003) and Ho (2009). They derive the profits that a provider can make by joining a network. Working with public contracts, Capps et al. (2003) find that all providers are contracted in equilibrium (see Section 3 below). This is not the case with private contracts. In Bardey and Rochet (2010), a health insurer operates in a two sided market: contracting providers upstream and selling insurance downstream to heterogeneous consumers. The equilibrium network is determined by two effects. The demand effect: insured value bigger networks. The adverse selection effect: bigger

[2] Note that the contracting here is between two parties that want to trade. In other words, the contracting is not about delegation and commitment as analyzed in papers like Kockesen (2007).

[3] This captures a rich set of non-linear contracts, including quantity restrictions and cross subsidies between different treatments (we only consider one treatment), used in practice. Moreover, (Laffont and Tirole, 1993, proposition 1.4) derive conditions under which the optimal contract in a model with asymmetric information can be implemented using two-part tariffs.

[4] This is the opposite of the problem in Hart and Tirole (1990). There, retailers pay a manufacturer for a profitable opportunity and worry that they may not get enough customers; here providers fear that they get too many patients.

[5] We do not analyze the choice between different bargaining procedures between insurers and providers. For an analysis of the choice between publicly posted prices and Nash bargaining between providers and insurers, see Barros and Martinez-Giralt (2008).

[6] Unless other constraints are introduced like the limited-liability constraint in Ma and Riordan (2002).

networks are relatively more attractive for high risk types. If the latter effect dominates, narrow networks are more profitable. As we focus on private contracting, our set-up is simpler with homogeneous providers and homogeneous consumers, constant treatment costs and no physician effort choice.

Gaynor et al. (2015) cite evidence suggesting that bigger insurers (e.g. due to a merger) manage to bargain lower prices with providers. This aspect of insurer size on bargained prices is orthogonal to the problem that we analyze for three reasons. First, we focus on symmetric insurers; that is, one insurer is not bigger than another due to mergers or efficiency differences. Hence, this effect does not show up in our symmetric equilibrium. Second, we focus in this paper on health care utilization and costs; not on bargained prices. Third, the evidence cited at the start of this introduction suggests that this effect of insurer size on bargained prices is not big enough for the following "virtuous circle": a wider provider network attracts more customers to an insurer, this increased insurer size improves the insurer's bargaining power to such an extent that health care costs fall for the insurer. As we noted at the start, a wider network is associated with higher costs; not lower costs. In other words, the bargaining effect of bigger insurer size is dominated by other (moral hazard) effects leading to higher costs. These latter effects are analyzed in this paper.

Finally, this paper is related to the industrial organization literature on private contracting. Papers in this literature specify agents' beliefs about contracts that are payoff relevant to them but which they do not observe. So-called passive beliefs are imposed by papers like Hart and Tirole (1990) and Segal (1999). With passive beliefs, a provider receiving a deviating contract from an insurer, still believes that other providers received the equilibrium contracts. We show that with passive beliefs, only fee-for-service contracts are used and no capitation ($t = 0$). As a technical innovation, we do not specify provider beliefs, but characterize the set of contracts that make it incentive compatible for the insurer to truthfully reveal to providers the relevant details of all contracts offered. We assume that providers are only willing to accept contracts that truthfully reveal how many patients they can expect in the coming period.

The rest of this paper is organized as follows. The next section introduces the model. This model abstracts away from many institutional features. It applies to contexts where profit maximizing insurers and providers bargain over treatment prices and where physicians' treatment decisions react to financial incentives. This can be the relatively unregulated private market in the US but also the more regulated Dutch market where health insurance is mandatory but sold by private insurance companies. In order to set the stage for private contracting, we consider five simple models with public contracting. These models can explain a relation between network size and health care expenditure, but not health care costs and utilization. Section 4 introduces private contracts. It shows the trade-off between supply-side cost sharing and provider profits. As the network grows, the optimal fee-for-service increases and the capitation fee falls. For a big enough network (think of indemnity insurance; see footnote[1]), there is no capitation fee at all. Section 5 characterizes the equilibrium in the insurance market. We conclude with policy implications. Proofs of results can be found in the appendix.

2. Model

Consider a model with risk averse consumers (mass one) buying insurance at premium $\sigma \geq 0$ with out-of-pocket payment $\gamma \geq 0$ (demand side cost sharing) in case a patient needs treatment. A patient needs at most one treatment per period. Treatment is provided by homogeneous risk neutral providers with cost $c > 0$ per treatment. Risk neutral insurers pay providers using two-part tariffs with capitation fee t and fee-for-service p. That is, the provider receives p each time she treats a patient, while the fixed fee t is paid once; say at the beginning of the period. The use of such two-part tariffs is common in the health care context (see, for instance, Christianson and Conrad, 2011; Gaynor et al., 2015, for overviews).

As we will see, to reduce over-consumption of health care services, the insurer wants to pay a fee-for-service which is less than the cost of treatment, $p < c$ (supply side cost sharing). The provider is compensated for this loss with $t > 0$. If the insured can only enroll with one provider, this provider receives the capitation fee per enrolled customer of the insurer. This is often how it works with a primary physician or family doctor. But requiring insured to enroll upfront with each possible specialist that may be needed in the coming period is not practical. In this case, people choose their specialist from the insurer's network once they need one. Hence, the capitation fee for provider P_i needs to take into account the probability that an insured falls ill and chooses P_i. One can think of the capitation fee t as a subscription fee: it gives the insured the right to be treated by the provider. One can think of an insurer paying t to the provider for each of its insured.

We follow Ma and McGuire (1997) and Bardey and Lesur (2006) in assuming that

$$p \geq 0 \tag{1}$$

Indeed, with a negative fee-for-service $p < 0$ and demand side cost sharing $\gamma \geq 0$, the patient and physician are better off not reporting the treatment to the insurer.

Note that compared to a pure capitation system ($p = 0$), a positive fee-for-service $p > 0$ can be seen as ex post insurance for the provider in case there is an unexpected increase in treatments provided.

With probability $\theta \in \langle 0, 1\rangle$ the agent falls ill. We do not consider adverse selection issues: θ is the same for all agents. To create an elastic demand for treatment, we assume that the value of treatment ν depends on the condition of the patient. The physician observes ν. We assume that $\nu \in [0, \bar{\nu}]$ is drawn from a distribution with cumulative distribution function F and density function f with $f(\nu) > 0$ for each $\nu \in [0, \bar{\nu}]$.

Taking into account the relation between physician and patient (Arrow, 1963; Ma and Riordan, 2002), we assume that they determine together whether the patient receives treatment or not. The three relevant parameters are (for given c): value ν of treatment for the patient (which the patient may or may not know), co-payment γ that patient needs to pay and fee-for-service p that provider receives from the insurer. We do not want to make very specific assumptions here. We assume that there exists a continuously differentiable function $\nu(p, \gamma)$ such that a patient receives treatment if and only if $\nu > \nu(p, \gamma)$.

In words, patients with high enough "severity" (value of treatment) ν are treated. The lower bound of who is treated is affected by financial incentives. Socially efficient treatment follows if $\nu(p, \gamma) = c$: a patient is treated if and only if the value of treatment ν exceeds the cost of treatment c. If $\nu(p, \gamma) > (<) c$ we say that there is under-(over-)treatment.

We make the following assumptions on the derivatives of ν with respect to p and γ:

$$\nu_p(p, \gamma) \leq 0, \quad \nu_\gamma(p, \gamma) \geq 0 \tag{2}$$

As the provider receives a higher compensation from the insurer for treatment, she is more willing to treat a patient (threshold decreasing in p); as the patient faces a higher co-payment, he is less keen to be treated (threshold increasing in γ). Both effects are well documented in the health economics literature. See Christianson and Conrad (2011) for an overview of the effects of supply side cost sharing and McGuire (2011) for an overview of demand side cost sharing found in the empirical literature.

Note that with $p < c$ it is still possible that the provider treats the patient. First, she is compensated for this via the capitation fee. Second, the patient may demand it and the physician gives in (though she is less likely to give in, the lower p is). Third, it may be obvious that the patient needs the treatment and the provider may fear legal consequences if the (loss making) treatment is withheld.

What we have in mind is that the patient and physician come to some sort of agreement on whether to treat or not. As treatment becomes financially more attractive for the physician, she is more likely to suggest it. As it becomes more expensive for the patient, he may be more reluctant to undergo treatment (Aron-Dine et al., 2013). One way to model this is to assume that physician and patient jointly have the following objective function

$$\beta(p - c) + (1 - \beta)(v - \gamma) \tag{3}$$

where $\beta \in [0, 1]$ captures the physician's bargaining power vis-a-vis the patient. The assumption here is that both the provider's and the patient's payoffs are linear in income and treatment value. This equation gives the value of treatment which is compared to the value of no treatment (normalized at) 0. The physician treats the patient if and only if (3) is positive; i.e. $v(p, \gamma) = \gamma + \beta/(1 - \beta)(c - p)$. This simple set-up is rich enough to allow for the following cases. Efficient collusion $\left(\beta = \frac{1}{2}\right)$ between physician and patient: patient is treated if $v + p - \gamma - c \geq 0$. Physician induced demand $\left(\beta > \frac{1}{2}\right)$: patient is treated because it is profitable for the physician even though patient's welfare may decrease ($v(p, \gamma) < \gamma$ if $p - c > 0$).

We normalize the function v to rule out less interesting cases; we assume

$$v(0, 0) > c \quad \text{and} \quad v(c, 0) < c \tag{4}$$

Cost c is high enough that there is under-treatment in case the physician receives zero fee-for-service ($p = 0$) and there is over-treatment if the physician is fully reimbursed ($p = c$); in both cases the patient faces no co-payment. In the former case, the physician is reluctant to give expensive treatment while this is not reimbursed to her at all ($p = 0$), in the latter, the patient – facing no costs – demands treatment while for the physician providing such treatment does not cost anything ($p = c$).

Hence, we rule out the case where there is either over-treatment or efficient treatment with $p = 0$. In this case, the optimal $p = 0$ and the analysis is rather trivial. Similarly, we rule out either under-treatment or efficient treatment at $p = c$. Again the optimal solution is straightforward in this case.

As we focus on the contracting between insurers and providers, we work with a simple utility function that is linear in treatment value and costs. We capture the agent's risk aversion by an additively separable dis-utility function $\delta(p, \gamma)$; think of this as the variance term in the agent's utility function. We normalize,[7] such that dis-utility for the agent equals $\theta\delta(p, \gamma)$ and make the following assumptions on δ. First, $\delta(p, 0) = \delta_p(p, 0) = 0$: as $\gamma = 0$ there is no dis-utility of financial risk because the agent does not face an out-of-pocket payment (for any value of p). Further, $\delta_\gamma(p, \gamma) > 0$ at $\gamma > 0$: an increase in γ raises the dis-utility as the agent needs to pay more in case he needs treatment; in this sense the risk increases with γ.[8] Finally, $\delta_p(p, \gamma) \geq 0$: as p increases, patient is more likely to be treated (and has to pay γ) because $v_p(p, \gamma) \leq 0$.

If the agent does not buy insurance, he goes to a provider in case he needs treatment. We assume that there is price competition on the uninsured market and treatment is a homogeneous good. Hence, the uninsured price equals $p^u = c$ and uninsured utility is given by

$$V^u = \theta \int_{v(c,c)} (v - c)f(v)dv - \theta\delta(c, c) \tag{5}$$

where we integrate over all states of the world where the patient gets treatment ($v > v(c, c)$). Without insurance, patient pays $\gamma^u = c$ for treatment and provider receives $p^u = c$ (from the patient) for the treatment. Hence the threshold treatment is given by $v(c, c)$. Sometimes, it is assumed that without insurance, efficient treatment choices are made. This would imply $v(c, c) = c$; although we allow for this, we do not impose it.

If the agent buys insurance at premium $\sigma \geq 0$, his utility equals:

$$V^i = \theta \int_{v(p,\gamma)} (v - \gamma)f(v)dv - \sigma - \theta\delta(p, \gamma) \tag{6}$$

Note that a high value of $v(p, \gamma)$ implies a high risk of not being treated. This, in turn, leads to a low value of insurance. In other words, insurance with high $v(p, \gamma)$ is not very valuable for the consumer.

An agent buys insurance if and only if $V^i \geq V^u$. Note that the values V^u, V^i capture the increase in patient's utility due to treatment, in case he is treated. If a patient falls ill, this will tend to reduce utility directly (dis-utility of not feeling fit). But this is not relevant for the value of treatment (and hence the expected value of being (un)insured) which is determined by the increase in utility due to treatment in the model here. If you are not treated, you forgo treatment's utility increase.

We exclude the case where a patient is paid for undergoing treatment (i.e. we exclude $\gamma < 0$). We are not aware such contracts are used in reality. Moreover, this would create serious moral hazard issues on the patient's side. Further, without insurance the patient pays $p^u = c$, hence we have with insurance that

$$\gamma \in [0, c\rangle \tag{7}$$

Consider the case with one provider P_1 and one insurer I_a. P_1's profits with contract p, t equal

$$\pi_1 = H(p, \gamma)(p - c) + t \tag{8}$$

where H denotes the probability that a patient is treated:

$$H(p, \gamma) = \theta(1 - F(v(p, \gamma))) \tag{9}$$

with – from Eq. (2) – $H_\gamma \leq 0$, $H_p \geq 0$. As explained above, a physician's treatment choice can be affected by patient utility as well. But we do assume that providers when bargaining with insurers only accept contracts that lead to non-negative profits. In this sense, providers focus on financial viability taking into account how their interactions with patients affect their treatment choices.

Insurer I_a's profits equal

$$\pi_a = \sigma - H(p, \gamma)(p - \gamma) - t \tag{10}$$

As a benchmark, note that social welfare is maximized by implementing efficient care consumption $v(p, \gamma) = c$ while minimizing $\delta(p, \gamma)$. Eq. (4) implies that there exists $p^* \in \langle 0, c\rangle$ such that

$$v(p^*, 0) = c \tag{11}$$

and $\delta(p^*, 0) = 0$. To make sure that provider's expected profits are non-negative, capitation fee needs to be equal to at least $t = H(p^*, 0)(c - p^*)$. Such a contract implements the first best outcome. Only supply-side cost sharing is used ($p^* < c$). Because the agent is risk averse, $\gamma > 0$ would be inefficient and thus there is no demand-side cost sharing in first best.

[7] In terms of notation, it is convenient if a number of terms are multiplied by θ; hence we multiply the function $\delta(.)$ by θ as well. This just normalizes the function $\delta(.)$. Since we do not do comparative statics with respect to θ, this normalization is without loss of generality.

[8] In fact, there is also another effect: as γ increases, the probability that the patient gets treatment decreases because $v_\gamma \geq 0$. If relevant, we assume that the direct effect of γ dominates this indirect effect.

With public contracts, the first best outcome can be implemented also with competing risk neutral providers $P_1, \ldots, P_N$. Suppose the insurer contracts n of these providers, then $p = p^*$, $t = H(p^*, 0)(c - p^*)/n$ leads to non-negative expected profits for providers. As providers are homogeneous, the probability that a patient visits P_i – conditional on being ill - equals $1/n$. Hence, with public contracts, moral hazard in health care consumption can be solved by supply side cost sharing (only). This leaves the question why moral hazard in health care is still an issue.[9] Further, health care costs do not vary with the size of the network n and AWP laws have no effect on the outcome.

In Section 4, we consider the effects of private contracting. To set the stage, we briefly consider arguments why – with public contracts – the size of the network can affect health care expenditures but not utilization and costs.

3. Public contracts

In the introduction we suggested that there is evidence for the following two related findings: (i) as the size of an insurer's network increases, health care costs increase and (ii) AWP laws tend to raise health care costs; where AWP laws "require managed care plans to accept any qualified provider who is willing to accept the terms and conditions of a managed care plan" (Hellinger, 1995, pp. 297).

There are a number of arguments in the literature why broader networks lead to higher health care expenditure. We provide simple formalizations of these arguments using public contracts. This shows that these arguments cannot explain the effect of (i) and (ii) above on health care utilization and costs. Additional assumptions are needed to get an effect on utilization and costs. For instance, a combination of arguments below together with public contracts may explain the findings mentioned in the introduction. The assumption we introduce in the next section is private contracting, which indeed is prevalent in health care markets.

3.1. Threat to exclude

The first argument explains the low expenditures of narrow networks by the threat to exclude. This threat enhances the insurer's bargaining power, leading providers to lower their prices. The bargaining effect is clear, but this does not imply that equilibrium networks are narrow nor that utilization is affected. We illustrate this using the model above.

Insurer I_a offers publicly each of the N providers a contract with

$$p^*, t^* = H(p^*, 0)(c - p^*)/N \tag{12}$$

where $p^* < c$ is defined in (11). Each provider is willing to accept this contract and all providers are contracted in equilibrium. Similarly, if providers make the offers to the insurer (bidding game),[10] they compete the price down to the same contract (p^*, t^*). All these contracts are accepted by the insurer.

According to this reasoning, there is no relation between the size of the network $n = N$ and either health care expenditure or costs. There can be an effect on price if I is *under the obligation* to contract with all N providers. Then each provider can claim part of the rent earned by the insurer in the bidding game, because the insurer cannot reject offers with $t > t^*$.[11] Utilization and costs are not affected because the efficient contract remains optimal. But – even with AWP laws – an insurer is not under the obligation to contract all providers nor to accept any contract offer that a provider makes.

Summarizing, the only thing that is needed, is the threat to exclude a provider. In the equilibrium of such a game, however, no provider is necessarily excluded. I offers insurance that maximizes consumer's value: $p = p^*$, $\gamma = 0$.

3.2. Shifting volume

Another argument in favor of narrow networks is that shifting patient volume to a small number of providers leads to lower costs. To model this, we need increasing returns to scale (IRS). Consider the set up above with two providers. Provider i's cost per treatment is denoted $c(x_i)$ where x_i denotes number of patients treated by P_i and $c'(x_i) < 0$ captures IRS. Let X^* denote the total number of patients. To simplify the exposition, assume X^* is exogenous here.[12] We have $x_1 + x_2 = X^*$. One option for I is to offer P_1 a contract with $p = 0$, $t = X^* c(X^*)$. As $c'(x) < 0$, it is optimal for I to deal with only one provider. So this model explains why a network with only one provider leads to lower costs than a network with both providers.

However, this argument is not completely convincing for two reasons. First, it is not clear that in this model AWP laws raise health costs. To see why they may have no effect, consider the case where I offers a menu with two contracts: one contract $p = 0$, $t = X^* c(X^*)/2$ for the case of a network with 2 providers and $p = 0$, $t = X^* c(X^*)$ for a network with one provider. If one provider accepts the latter contract, the other provider has no incentive to join the network with the former contract. Hence, AWP laws do not affect the outcome in this set up. Indeed, the insurer does not exclude a provider; the provider is not willing to join the network.

Second, the main reason why IRS is not a very convincing explanation for a narrow provider network is that the optimal hospital size is quite modest. As pointed out by Haas-Wilson (2003, pp. 147) for hospitals "most scale economies appear to be exhausted at relatively low levels of output". Posnett (1999, pp. 1063) summarizing empirical studies notes that "research does not support any general presumption that larger hospitals benefit from economies of scale". In fact, dis-economies of scale can set in, making it optimal to spread an insurer's patient population over a number of hospitals.

3.3. Taste for variety

Horizontal product differentiation is also used as an explanation for why bigger networks lead to higher health care expenditure – not costs (see, for instance, Gal-Or, 1997). To illustrate this, assume that there is one insurer and two providers. The providers are located on a Hotelling beach of length 1 with provider 1 on position 0 and provider 2 on position 1. Let t denote the travel cost over the beach. When an agent buys insurance, he does not know yet where he ends up on the beach once he needs treatment; assume that each location between 0 and 1 is equally likely (uniform distribution for the agent). Intuitively, each provider may specialize in certain treatments and when buying insurance, the agent does not know yet which treatment he needs.[13]

---

9 One possible explanation is asymmetric information (about provider costs and/or quality). This is explored in Boone and Douven (2014).

10 The case where a monopoly insurer makes the offers to providers is then called an offer game (Segal and Whinston, 2003).

11 If providers claim more then the rents earned by I, the insurer closes down to earn its outside option.

12 More generally, in the set up above, let p^* denote a solution to $c(H(p^*, 0)) = v(p^*, 0)$, then $X^* = H(p^*, 0)$.

13 Another form of provider variety is noise in diagnosis. Assume, as above, that a patient is treated if his "severity" v exceeds a benchmark value. But the observed severity equals the "true" severity plus some noise. By shopping around, a patient can try to find a provider P_i who diagnoses v_i in excess of the benchmark. This could also explain why bigger networks have higher utilization. Two observations make this argument less compelling. First, even if there is one provider in the network, the patient can return to get a new diagnose (e.g. claiming things got worse). Second, if a patient was denied treatment by 3 providers, the 4th provider giving the treatment has something to explain to the insurer.

If the insurer contracts only one provider (exclusive contract), expected value of insurance for the agent equals $u_e = \theta\left(v - \frac{1}{2}t\right)$ where v denotes the value of treatment once it is needed and $\frac{1}{2}t$ the expected travel cost. If both providers are contracted (common outcome), we have $u_c = \theta(v - \frac{1}{4}t)$. Hence, a monopoly insurer charges a higher premium if both providers are contracted: $\sigma_c = u_c > u_e = \sigma_e$. But the insurer offers the same prices to providers in both cases: p^* to ensure efficient treatment decisions (and a capitation fee to keep expected provider profits non-negative). Hence, health care costs in this case are unaffected, though a monopolist insurer charges a higher premium in case of a bigger network.

If providers make the offers, they can demand a higher capitation fee to appropriate part of the surplus associated with the common outcome.[14]

Although the distribution of rents is different depending on who makes the (public) offers, welfare and efficiency are unaffected. From a welfare point of view, capitation fees are transfers between parties without affecting treatment decisions or efficiency. Hence, this type of model cannot address efficiency concerns that smaller networks tend to reduce access to physicians and decrease the number of treatments (Terhune, 2013).

3.4. Heterogeneous providers or agents

Our model focuses on homogeneous providers and symmetric agents. Cost heterogeneity can also explain why bigger networks tend to have higher costs. However, it cannot explain well the two observations at the start of this section.

First, consider the case where providers have different costs and treatment decisions are exogenous. A narrow network – that only contracts the most efficient providers – has lower costs than a network that also contracts less efficient providers. To formalize this idea, some form of (horizontal) provider differentiation is needed (like the Hotelling set up above); otherwise, what is the value of contracting inefficient providers?

But in such a model it is hard to understand why AWP laws affect health care costs. Suppose the efficient providers treat patients at cost c per treatment, while less efficient providers have costs $c' > c$. Leaving capitation aside,[15] an insurer can offer a contract with fee-for-service $p = c$ (or a slightly higher p). Any provider willing to treat at this price can accept the contract. Inefficient providers will not accept such a contract; AWP laws do not force insurers to offer contracts with $p \geq c'$. Further, this line of argument suggests that all narrow networks contract the same (efficient) providers. Although there is some overlap, the lack of overlap in a number of areas (Coe et al., 2013, pp. 9) makes this argument less convincing.

Second, consider agents with different expected health care costs and adverse selection. If providers are (perceived to be) differentiated in utility space, broader networks are more attractive to insured than narrow networks. It seems reasonable to assume that this preference is stronger for people with higher expected health care costs. Due to this effect, a cross section of insurers where some offer narrow and others broad networks tends to show that the broader networks have higher costs per insured (Cutler and Reber, 1998; Coe et al., 2013; Bardey and Rochet, 2010). But this argument is not directly convincing to explain why states with AWP laws tend to have higher health care costs. Admittedly, there can be some endogeneity here, but it is not clear that this can explain cost differences of 3%; see the analysis of Klick and Wright (2014) using state fixed effects. Further, adverse selection cannot explain the change in costs when a country moves from indemnity insurance, via managed care with narrow networks to a situation with broader networks (Cutler, 2004; Lesser et al., 2003).

3.5. Risk averse providers

Capitation contracts transfer risk from the insurer to the provider. We assume that providers are risk neutral; the capitation fee only needs to cover the expected loss from treatment. However, if providers are risk averse, insurers need to compensate them for this risk. Can this explain the relation between network size and health care costs?

The risk that a provider faces is the loss $c - p > 0$ per treatment and the insurer has to pay a risk premium. To reduce the risk premium, the insurer offers $p > p^*$. This increases health care utilization and costs (compared to a situation with risk neutral providers). The effect of an increase in network size is, however, ambiguous. As network size n goes up, more providers need to be paid a risk premium; this tends to raise p and costs. However, as n increases, the probability that one provider incurs the loss $c - p$ decreases: the risk is spread over more providers. If the latter effect dominates, p and costs tend to fall with n.

Further, as hospitals perform most operations hundreds of times a year, the law of large numbers would suggest that there is not that much risk at the provider level. Hence for routine treatments, the risk premium for providers is not really an issue and $p = p^*$. With private contracts, the risk that providers face is not stochastic but strategic (influenced by the insurer).

This section briefly presented explanations for the relation between network size and health care expenditure with public contracting. We argued that these explanations do not explain an effect of AWP laws on utilization and costs.

4. Private contracts

The model that we use has two ingredients. First, contracts between insurers and providers are private (i.e. not publicly observable). Second, the number of patients treated by provider P_i depends on the fee-for-service p offered to other providers in the insurer's network. Initially, we capture the latter effect by assuming that insurers can guide patients to certain providers within their network.

We are interested here in implicit mechanisms by which insurers steer patients to providers. Explicit mechanisms to steer patients are excluding providers from the network and charging patients different co-payments for different providers. These mechanisms are explicit because they need to be specified in a consumer's insurance contract. Hence, these mechanisms are contractible for providers as well. Instead, we focus on ways to steer patients which are not verifiable for providers. Such implicit mechanisms include advising patients directly when they need to choose a provider. As insurers know how other customers fared with certain hospitals and physicians, they have relevant information for patients. Patients can contact their insurer to ask about this. Or the information can be presented on a web-site or in an app (De La Merced, 2014; Scott, 2011). Presenting providers in the network on a website in a certain order will affect patients' choices. Finally, insurers can influence primary physicians to steer patients to certain providers and not to others (Ho and Pakes, 2013; Liu, 2013; Kirk, 2014). Such mechanisms are implicit as they cannot be verified by hospitals.

[14] Using Bernheim and Whinston (1998), providers set p^* to maximize efficiency in both the common and exclusive case. Competition with exclusive contracts leads to zero rents for providers (as in the offer game we consider). With common contracts, each provider claims her contribution to the surplus as rent (here the reduction in expected travel cost $\theta\frac{1}{4}t$). Gal-Or (1997) uses Nash bargaining to model the distribution of rents.

[15] With capitation and endogenous treatment decisions, insurer offers $p = p^*$ and capitation that covers the difference $c - p^*$. Providers with costs $c' > c$ do not accept this contract.

Although, (implicit) steering is plausible and simplifies the exposition, Section 4.6 shows that this assumption is not necessary to get that P_i's profits depend on p_j received by provider P_j ($j \neq i$).

Hart and Tirole (1990) introduced private contracts in the context of an upstream monopolist with downstream retailers. In their model, private contracting is a contracting inefficiency: the monopolist cannot commit to public contracts. If it could commit to such contracts, it would be better off. This is different in our model. In fact, with Bertrand competition, insurer profits are always zero (Section 5). But provider profits are positive with private contracts where they would be zero with public contracts. Hence, providers actually have a stake in defending the confidentiality clauses in their contracts.

This section explains how private contracting affects the relation between network size and costs. Section 6 comes back to AWP laws.

4.1. Capitation and implicit steering

In addition to the elements in Section 2, we have the following model in mind. Insurers $I_a, \ldots, I_m$ simultaneously and independently offer providers $P_1, \ldots, P_N$ contracts[16] with a fee-for-service p and capitation t.[17] Without observing offers that other providers received, providers simultaneously and independently decide which offers to accept. These acceptance decisions determine each insurer I_j's network size n_j. Each insurer sells an insurance contract specifying its network, co-payment γ and premium σ.[18] Since providers are homogeneous, an insurer's network is characterized by its size. As n and γ are specified in the insured's contract, these parameters are contractible for providers as well.[19] Hence, the contracts that providers receive, are conditional on n and γ. As the insurer's choice of γ is payoff relevant to providers ($H_\gamma \leq 0$), it is specified in its contracts with providers. This γ is then also used in the insurance contract sold to consumers. Because the probability of being treated depends on the fee-for-service, p is payoff relevant for consumers. However, p is not observed by them due to the private contracting between providers and insurers. Indeed, few people know the prices specified in providers' and insurers' contracts. But, as we show below, an insurer's network size signals p. Based on σ, γ, n, an agent decides whether to buy insurance and from whom. Once an insurer knows its number of customers, it pays t upfront for each customer to the providers in its network.

Then consumers fall ill and need to go to a provider. As described above, we assume that insurers can steer consumers to providers. As providers are homogeneous, consumers do not object to this.[20] To keep things simple, we assume that patients do not incur (travel) costs to visit a provider. If a provider does not treat the patient, he can visit another provider in his insurer's network (if there is one), etc.[21] If the patient is treated by a provider, the patient pays γ to the insurer and the provider receives p from the insurer.

Summarizing, timing of the game is as follows:

1. $I_a, \ldots, I_m$ simultaneously, independently and privately make offers of the form (p, t, γ) to $P_1, \ldots, P_N$;
2. providers simultaneously and independently accept/reject offers; this gives I_j's network size n_j;
3. each I_j announces premium σ_j, network size n_j and co-payment γ_j to consumers;
4. consumers decide whether to buy insurance and from which insurer;
5. I_j pays contracted t_j to providers for each insured customer;
6. patients fall ill;
7. I_j guides/steers patients to providers;
8. provider P_i treating patient receives contracted p_{ij} from I_j; patient pays γ_j to I_j.

Because of assumption (4), the insurer wants to set $p < c$ to reduce over-consumption of health care. To compensate for the loss of treating a patient, providers are paid a capitation fee. When evaluating an offer (p, t), a provider needs to take into account the probability that she will treat an insurer's customer.

This is straightforward with public contracts. Consider an insurer who contracts two providers $P_{1,2}$ with $p_1 < p_2 < c$. Then $t_1 = H(p_1, \gamma)(c - p_1)$ as P_1 understands that all patients from this insurer are first steered to her. Further, $t_2 = (H(p_2, \gamma) - H(p_1, \gamma))(c - p_2)$ since P_2 only treats patients that are not treated by P_1.

However, this does not work with private contracts. In particular, whereas the contract with provider P_i can specify the size of the network n and co-payment γ (both are verifiable information), P_i does not know how p_i relates to the fee-for-service offered to other providers in the insurer's network. If p_i is the lowest fee offered in the network, P_i should expect to treat $H(p_i, \gamma)$ patients; if there are providers P_j in the network with $p_j < p_i$, P_i treats fewer patients than $H(p_i, \gamma)$.[22]

As an example, consider I_a offering $P_{1,2}$ prices $p_1 \leq p_2 \leq c$ and $t_1 = H(p_1, \gamma)(c - p_1)$ and $t_2 = (H(p_2, \gamma) - H(p_1, \gamma))(c - p_2)$.[23] If P_2 cannot observe/contract on p_1 and t_1, does she accept the contract above? We argue that she does not. If she would accept, I_a has an incentive to deviate and offer P_1 the same contract p_2, t_2. For p_2 close enough to p_1, paying for the probability of treatment $H(p_2, \gamma) - H(p_1, \gamma)$ – even if it is paid twice – is less than paying for the probability $H(p_1, \gamma)$.

Summarizing, with private contracts an insurer makes each provider P_i an offer (p_i, t_i); where the offer is conditional on public information about the insurer's network size n and co-payment γ. The insurer has payoff relevant information (prices p_j for $j \neq i$) that P_i cannot observe but needs to know to evaluate the expected profits associated with (p_i, t_i). There are two ways to proceed. First, given the offer (p_i, t_i), P_i forms beliefs about the other offers (p_j, t_j). Second, the insurer truthfully reveals its private information. We use the latter route and come back to beliefs in Section 4.5.

The next subsection characterizes the menu of contracts that leads to truthful revelation by insurers. We assume that providers

[16] There are games where the parties making the take-it-or-leave-it offers have all the bargaining power and capture the rents. But this is not always the case. To illustrate, in a Bertrand competition game with homogeneous goods and constant marginal costs, the firms make the offers and earn zero profits. Something similar happens here as well: the insurers make the offers but earn zero rents under Bertrand competition on the health insurance market. Instead, the providers make positive profits.

[17] Offers made by insurer I_i to provider P_j can be written as (p_{ij}, t_{ij}). To ease notation, we drop the subscript ij if this does not cause confusion. Notation (p, t) should not be interpreted as symmetric offers.

[18] Since consumers are identical, there is no need for an insurer to offer more than one contract.

[19] In principle, we could consider co-payments that vary with provider. But because providers are homogeneous and γ is contractible, there is no reason to do so.

[20] See Boone and Schottmüller (2019) for an analysis of the case where providers differ both in costs and in quality.

[21] In equilibrium this happens only if providers receive different p's from a patient's insurer: see below.

[22] Note that the nature of this contracting problem between insurers and providers differs depending on whether $p > c$ or $p < c$. The relevant problem in our context is $p < c$. With $p > c$, treating the patient becomes a profitable opportunity and the provider is willing to pay the insurer for this opportunity ($t < 0$). This is the problem analyzed in Hart and Tirole (1990).

[23] To ease notation, we will not always explicitly acknowledge that $p_1 = p_2$ is also possible (but we do not exclude this possibility). With $p_1 = p_2$ the lowest price, we think of H and t_i as satisfying $t_1 = t_2 = \frac{1}{2}H(p_1, \gamma)$.

only accept contracts where they know the number of patients that they can expect to treat. From the menu of truthful contracts, we derive the optimal contracts for insurers. Although final consumers cannot observe the private contracts between insurers and providers, they can infer these contracts from the insurer's network size. After seeing network size and demand side cost sharing, consumers use Bayesian updating to infer their probability of treatment with a certain insurer. This determines the premium they are willing to pay for health insurance. Section 5 then derives the equilibrium in the health insurance market under perfect and imperfect competition between insurers.

4.2. Truthful revelation

We take a mechanism design approach and characterize the set of contracts that make it incentive compatible (IC) for an insurer to reveal its private information truthfully to providers. As the insured do not observe the offers (p_i, t_i), the insurer's revenues (from premiums paid by the insured) cannot depend on p_i, t_i. For a given network size, IC refers to I's total costs C (only). That is, IC implies that there is not a deviating contract that strictly reduces I's expected costs; taking the (contractible) network size n as given.

Intuitively, I would like to offer a provider a contract with $p_i = 0$ and tell the provider that she should not expect any patients at all and hence $t_i = 0$. However, since $p_i = 0$ the insurer has an incentive to send its patients who need treatment to this provider precisely because treatment is free for the insurer ($p_i = 0$). Therefore, the provider should not believe that she will not treat any patients. As treatment is costly ($c > 0$) for the provider, this contract with $p_i = t_i = 0$ will be a loss making contract and the provider should reject the contract.

This is a clear cut case; but what if $p_i > 0$ but still below c and $t_i > 0$ but small? Should the provider accept such an offer based on the insurer's claim that she will not treat many patients? This will depend on the contracts offered to the other providers. If all the other providers have contracts with $p_j < p_i$, it is likely that P_i will not treat many patients as all the other providers offer cheaper treatment ($p_j < p_i$) for the insurer. That is, the insurer has an incentive to guide its patients to other providers than P_i. However, if $p_j > p_i$ for all other providers, the insurer has an incentive to get the patients to P_i making it unlikely that a small t_i covers the loss $p_i - c < 0$ per patient.

To model the set of contracts that P_i can accept because she can trust the contracts to avoid (expected) losses, we characterize the set of contracts where the insurer truthfully reveals the number of patients that a provider can expect to treat. If the number of patients is truthfully revealed, then the corresponding per capita fee t_i will (in expectation) cover the loss per treatment.

To characterize this set of acceptable contracts, we introduce some notation. Let x_i denote the true probability that a patient in the insurer's network is treated by P_i — given all of I's contracts — and $\hat{x}_i$ denotes I's message ("promise") to P_i of this probability. Given this message, I offers P_i a price per treatment $p_i \le c$ and a fixed fee $t_i = \hat{x}_i(c - p_i)$. We define the set of contracts (p, t) where x_i is truthfully revealed as

$$A_{\gamma,n} = \{(p, \hat{x}(c-p)) \in [0, c] \times \mathrm{IR}_+ | \hat{x} \ge x\} \quad (13)$$

where the set $A_{\gamma,n}$ is conditional on contractible information γ, n. Both the co-payment and the network size affect a provider's payoff. As the co-payment (γ) is higher, patients are less keen on treatment and the provider's expected treatment probability and treatment cost decrease. As the network is broader (higher n), the probability that a provider needs to treat a patient decreases (ceteris paribus). Since the network and the co-payment are contractible for the insured, they can also be contracted upon with providers.

Given that the treatment probability x is truthfully revealed, we assume that a provider is willing to accept each contract in $A_{\gamma,n}$. Accepting this contract will not lead to a loss (in expectation). Moreover, a contract that is not in the set $A_{\gamma,n}$ implies that $\hat{x}$ is not truthful. That is, the insurer suggests that fewer patients need to be treated than what will be realized in expectation. Such contract is likely to lead to a loss for the provider. Providers reject contracts that are not in the set $A_{\gamma,n}$.

Since a provider accepts contracts in $A_{\gamma,n}$ and rejects contracts that are not in this set, the optimal action of an insurer is to offer contracts in $A_{\gamma,n}$ with the lowest cost for the insurer. For a given p, that implies the contract with the lowest t that is still in this set of acceptable contracts. Hence, only contracts in $A_{\gamma,n}$ are offered and accepted. We now characterize this set of acceptable contracts.

Proposition 1. *For each $(p, t) \in A_{\gamma,n}$ we have that*

$$t \ge H(p, \gamma)(c - p) \quad (14)$$

In words, a provider who receives $(p, t) \in A_{\gamma,n}$ views the offer as if p is the lowest price offered to all n providers in the network: $\hat{x} = H(p, \gamma)$. It is as if the provider expects all patients that need treatment to come to her. If, instead, she would think that there is another provider with the same low price p, she would be willing to accept a contract with $t = \frac{1}{2}H(p, \gamma)(c - p)$. The proposition claims that such a contract should not be trusted by the provider as it is likely to lead to a loss if it is offered by an insurer. Any contract that is not in $A_{\gamma,n}$ cannot be trusted to be truthful. Assuming that the provider does not want to be lied to, she rejects contracts that are not in the set of acceptable/truthful contracts.

The provider should only accept contracts with a capitation fee that is high enough to cover her losses in case she is the only provider contracted by the insurer. That this is sufficient for a contract to be in $A_{\gamma,n}$ is clear: such a contract cannot lead to a loss (in expectation) for the provider. The proof shows that inequality (14) is also necessary for a contract to be acceptable to a provider (truthful for an insurer).

Note that transfers t in the set A are, in fact, independent of the size of the network n. This is the cost for the insurer of using capitation contracts in a network with competing providers. Although capitation contracts lower health care consumption by setting $p < c$, each provider worries that patients are steered towards her first. Hence, due to the strategic uncertainty regarding insurers the capitation fee can exceed the expected treatment loss for providers.

Although the reasoning above looks a bit "sterile" for providers to consider when bargaining with insurers, the underlying problem is, in fact, very concrete. In contrast to GP's, patients do not register in advance with hospitals. So when a hospital accepts a capitation fee (that is, a yearly fee that is independent of the number of patients or treatments), it does not really know what its (potential) patient population is. Hence, there is uncertainty about which patient will choose this hospital in case he needs, say, heart surgery. Part of this uncertainty is strategic: it is affected by the insurer both through the prices the insurer offers to other hospitals in the area and through the steering of patients by the insurer.

This strategic uncertainty leads a hospital to demand a rent from insurers to cover treatment losses ($p < c$): better safe than sorry. This "better safe than sorry" strategy can create strictly positive profits for providers. The provider only accepts contracts as if she is the only provider contracted by the insurer even if the insurer's network is, in fact, bigger ($n \ge 2$). The insurer pays a capitation fee as if all patients in the insurer population register with this provider. If a provider would be willing to accept contracts with a lower capitation fee (in case of $n \ge 2$), it opens the door for the insurer to offer contracts that lead to a loss in expectation for the provider.

At first sight then, the ability to (implicitly) steer the patient hurts the insurer. It raises the capitation fees that an insurer has

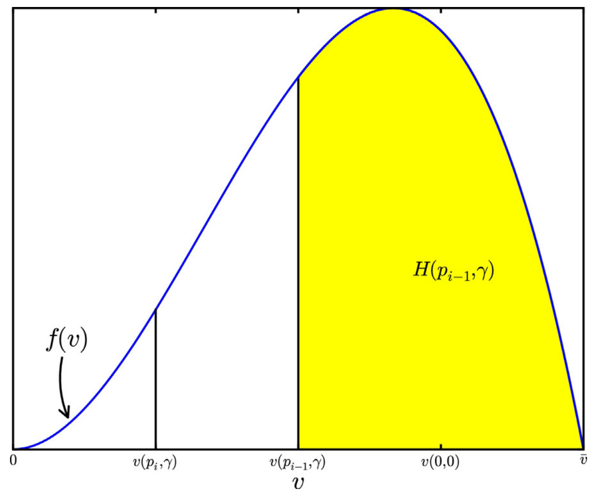

Fig. 1. Provider P_i makes a profit over $H(p_{i-1}, \gamma)$ patients not treated by her.

to pay. However, this is a blessing in disguise. First, note that by contracting only one provider this problem is resolved (recall that providers are homogeneous). Second, as we show shortly, the result in Proposition 1 allows the insurer to signal its fee-for-service to the consumer.

4.3. Minimizing costs

Given the set of acceptable contracts, in this section we characterize insurers' optimal contracts to providers.

Since p_i, t_i is not observed by final consumers, these contracts cannot affect an insurer's revenues. Therefore an insurer chooses its contracts with suppliers to minimize its costs:

$$C(n,\gamma)=\min_{p_i\le p_{i+1}} H(p_1,\gamma)(c-\gamma)+\sum_{i=2}^{n}[(H(p_i,\gamma) -H(p_{i-1},\gamma))(p_i-\gamma)+H(p_i,\gamma)(c-p_i)] \tag{15}$$

where we order providers such that $p_1 \le p_2 \le \ldots \le p_n$ and t_i is chosen such that Eq. (14) holds with equality (no reason for I to give more than this). Provider P_1 gets all patients to visit her first. She treats $H(p_1, \gamma)$ patients. Hence, I spends $H(p_1, \gamma)(p_1 - \gamma) + t_1 = H(p_1, \gamma)(c - \gamma)$ on this provider (in expected terms). If the patient is not treated by P_1, he will go for a "second opinion" and I steers him towards P_2. Provider P_i ($i \ge 2$) receives t_i and treats $H(p_i, \gamma) - H(p_{i-1}, \gamma)$ patients. I spends $(H(p_i, \gamma) - H(p_{i-1}, \gamma))(p_i - \gamma) + H(p_i, \gamma)(c - p_i)$ on this provider. This leads to a profit (per insured) for P_i equal to

$$\pi_i = H(p_{i-1},\gamma)(c-p_i) \tag{16}$$

P_i gets capitation fee based on treatment probability $H(p_i, \gamma)$, while the probability that she actually treats equals $H(p_i, \gamma) - H(p_{i-1}, \gamma)$. As illustrated in Fig. 1, $H(p_{i-1}, \gamma)$ of the patients are not treated by P_i and she makes a profit on these patients.

Hence, although providers are homogeneous and insurers make take-it-or-leave-it offers, providers make strictly positive profits if $n \ge 2$ and $p_n < c$. Define the sum of provider profits in a network of size n as

$$\Pi_P(\gamma,p_1,\ldots,p_n)=\sum_{i=1}^{n}\pi_i=\sum_{i=1}^{n-1}H(p_i,\gamma)(c-p_{i+1}) \tag{17}$$

It is routine to verify that (15) can be written as

$$C(n,\gamma)=\min_{p_i\le p_{i+1}} H(p_n,\gamma)(c-\gamma)+\Pi_P(\gamma,p_1,\ldots,p_n) \tag{18}$$

The first term is the expected cost of treating the patient. If the patient falls ill, the probability that he is treated (at all) equals $H(p_n, \gamma)$ with p_n the highest fee-for-service in the network. If expected provider profits would be zero, the first term would be I's treatment cost. The second term equals total providers' profits.

4.4. Optimal fee-for-service

With $n = 1$, provider profits equal 0 and costs are minimized by minimizing the treatment probability. For given γ, the profit maximizing fee-for-service $p = 0$. With $n \ge 2$, the trade-off faced by the insurer is the following. If it sets the highest $p_n = 0$, utilization is low but provider profits Π_P are high. Setting the lowest $p_1 = c$, leads to $\Pi_P = 0$ but a high probability of treatment.

Before characterizing the equilibrium in the insurance market, we need to characterize the effects of the contractible variables co-payment γ and network size n on insurer's costs C and highest price p_n. We assume that patients face no costs visiting providers (apart from the copayment when they are treated). Hence, for insured the relevant variable is the probability that they get treatment (at all) which is determined by p_n.

Proposition 2. *Costs $C(n, \gamma)$ are decreasing in γ and increasing in n for $p_n < c$.*

Highest price p_n is weakly increasing in n.

We find the following results. A higher co-payment γ leads to lower costs for the insurer and increasing the network size n leads to higher costs for the insurer. Further, increasing the network size also increases the optimal fee-for-service and hence the probability that a patient is treated in the network. The intuition is as follows.

Increasing the co-payment γ reduces costs directly (as patients pay a bigger contribution to the cost) and indirectly by reducing the probability of treatment ($H_\gamma \le 0$). As treatments become more expensive for the patient, he is less keen to accept a treatment offered by a provider.

Increasing network size n raises costs (unless $p_n = c$; then adding more providers does not affect costs as $t_n = 0$).[24] As more providers are contracted, low p_n becomes more expensive as a high capitation has to be paid to the contracted providers (see Proposition 1). To reduce this high capitation that needs to be paid to providers, the fee-for-service p_n tends to rise with n. Hence, we find that health care utilization and costs increase with network size. As n increases, the optimal fee-for-service increases and thus the probability of treatment increases because $dH/dn = H_p dp_n/dn \ge 0$.

This is our explanation for the observations in the introduction that bigger networks tend to go hand in hand with higher health costs. As the network grows, using capitation contracts with low fee-for-service becomes more expensive. Therefore, fee-for-service p increases with network size and thus health care utilization and costs increase with network size.

Note that $dH/dn \ge 0$ is line with many people's view that bigger networks tend to be more generous. As the network expands, supply side cost sharing becomes more expensive for the insurer. In a big network, a patient is more likely to receive treatment.

As shown in the proof of the proposition, the effect of γ on p_n is ambiguous. On the one hand, higher γ reduces the treatment cost $c - \gamma$; hence the insurer is willing to choose higher p_n and treat more patients. On the other hand, higher γ implies – ceteris paribus p_{n-1} – that P_{n-1} treats fewer patients and hence a lower profit has to be paid to P_n. This leads the insurer to choose lower p_n. Finally, there is the interaction effect $H_{p\gamma}$ on which we have not made any assumptions. Hence, we cannot sign $dp_n/d\gamma$ in general; neither do we need to sign it for our purposes.

[24] In fact, if $p_n = c$, we have $p_i = c$ for each provider i.

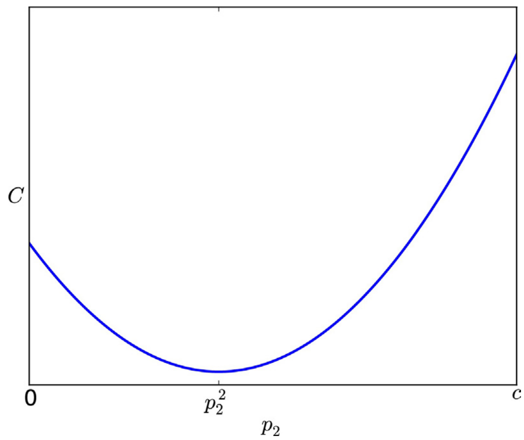

Fig. 2. Cost minimizing fee-for-service p_2 for provider P_2.

The following example illustrates the analysis above.

Example 1. Assume that $v \in [0, 2]$ with $f(v) = 1 - \frac{1}{2}v$. Patient is treated if and only if (3) is non-negative. There are two providers $P_{1,2}$. Consider the case with $\gamma = 0$. Then it is routine to verify that

$$H(p, 0) = \frac{\theta}{4}\left(\frac{\beta}{1-\beta}(c-p) - 2\right)^2 \tag{19}$$

If I contracts with P_1 only ($n = 1$), it is optimal to set $p_1^1 = 0$, $t_1^1 = H(0, 0)c$ – where the superscript denotes network size n. Health care costs equal $C(1, 0) = H(0, 0)c$.

If I contacts with both providers ($n = 2$), it is still optimal to set $p_1^2 = 0$, $t_1^2 = H(0, 0)c$. Optimal p_2^2 depends on parameter values. I's costs as a function of p_2 can be written as

$$C = H(0, 0)c + (H(p_2, 0) - H(0, 0))p_2 + H(p_2, 0)(c - p_2) \tag{20}$$

Fig. 2 draws C as a function of p_2 for parameter values: $\beta = 0.75$, $c = 0.2$, $\theta = 0.1$. Cost minimizing $p_2^2 \in [0, c]$ is an interior solution. As the size of the network increases, more patients are treated ($p_2^2 > p_1^1 = 0$) and health care costs equal $C(2, 0) > H(0, 0)c = C(1, 0)$.

4.5. Passive beliefs

In this paper, we focus on the set of offers where I truthfully reveals x_j to each P_j, in the sense of Eq. (13). An alternative that is often used in the literature on private contracting is to assume passive beliefs (see, for instance, Hart and Tirole, 1990; Segal, 1999). We argue that with passive beliefs it is not possible to use capitation contracts to reduce over-treatment.

With passive beliefs, a provider receiving a deviating offer believes that all other providers still received their equilibrium offer. Then there is only an equilibrium with $p_i = c$, $t_i = 0$ for all i. Suppose not, that is consider a Bayesian equilibrium where $p_1 \leq p_2 < c$ and both offers are accepted. Then a deviating offer with $\tilde{p}_2 = p_1 + \varepsilon$, $\tilde{p}_1 = p_2 + \varepsilon$ leads to $\tilde{t}_2 = (H(p_1 + \varepsilon, \gamma) - H(p_1, \gamma))(c - p_1) \approx 0 < t_1$ and $\tilde{t}_1 = (H(p_2 + \varepsilon, \gamma) - H(p_2, \gamma))(c - p_2) \approx 0 < t_2$. Hence such a deviation is always profitable for I. Therefore, with passive beliefs there are only fee-for-service contracts in equilibrium and no capitation fees: $p_i = c$, $t_i = 0$ for each i.[25]

Although contracts with $p_i = c$, $t_i = 0$ are in the set of acceptable contracts $A_{\gamma,n}$, they are not necessarily optimal in our mechanism design approach. Eq. (4) implies that $p = c$ leads to over-treatment. An insurer would like to introduce supply side cost sharing ($p < c$) to reduce over-treatment. The smaller an insurer's network, the lower the fee-for-service that it sets (Proposition 2).

4.6. No steering

We assume that patients have no preference for a particular provider and allow the insurer to guide them to a provider in the network. This simplifies notation, but is not essential for the results. To illustrate this, we sketch a model where a fraction $\alpha \in [0, 1]$ of insured chooses a provider on their own without consulting with their insurer.

Assume that there are two providers $P_{1,2}$ in insurer I's network. A fraction α of insured visits the provider that is closest to where they live; say, $\frac{1}{2}\alpha$ go to P_1 first and $\frac{1}{2}\alpha$ to P_2. If they are not treated by their chosen provider, they visit the other one.

Assume $p_1 < p_2$. How many patients are treated by each provider? First, consider P_1: she is visited by $\frac{1}{2}\alpha$ patients living close to her and $1 - \alpha$ patients guided by the insurer:

$$\frac{1}{2}\alpha H(p_1, \gamma) + (1 - \alpha)H(p_1, \gamma) = H(p_1, \gamma)\left(1 - \frac{1}{2}\alpha\right) \tag{21}$$

A fraction $\frac{1}{2}\alpha$ go to P_2 directly, whereas $1 - \frac{1}{2}\alpha$ go to P_1 first. Of these $1 - \frac{1}{2}\alpha$ patients, a fraction $H(p_1, \gamma)$ is treated by P_1 and $H(p_2, \gamma) - H(p_1, \gamma)$ is treated by P_2. Total number of patients treated by P_2 is given by

$$\frac{1}{2}\alpha H(p_2, \gamma) + \left(1 - \frac{1}{2}\alpha\right)(H(p_2, \gamma) - H(p_1, \gamma)) = H(p_2, \gamma) - \left(1 - \frac{1}{2}\alpha\right)H(p_1, \gamma) \tag{22}$$

That is, total number of patients treated by the network is $H(p_2, \gamma)$, $\left(1 - \frac{1}{2}\alpha\right)H(p_1, \gamma)$ of these are treated by P_1. For each $\alpha \in [0, 1]$, the probability that P_2 treats a patient depends on p_1 and we need the insurer to reveal p_1 truthfully to P_2. As α increases, the effect of p_1 on P_2's costs becomes smaller, but it does not disappear. Even with $\alpha = 1$, the patients not treated by P_1 will come to P_2 and hence P_2's profits depend on p_1.

5. Insurance market

In this section, we characterize the equilibrium on the health insurance market. Note that the effects in Proposition 2 are due to cost minimization; no assumptions are needed about market power in the health insurance market. To illustrate this, we first consider Bertrand competition between insurers. With Bertrand competition, no further distortions are introduced. Then we analyze the effects of market power in the insurance market.

5.1. Perfect competition

Consumers do not observe contracts (p, t) between insurers and providers. Hence, consumers' valuation of health insurance contracts cannot depend on these. Consumers do observe n, γ, σ and base their valuation of an insurance contract on these variables. Insurers that offer the same network size n and co-payment γ, offer homogeneous products.

As consumers are all the same, Bertrand competition between insurers leads them to offer the same contract n, γ at an insurance premium equal to $\sigma = C(n, \gamma)$. The probability that an insured patient is treated (at all) depends on the highest contracted fee-for-service p_n which we denote by $p(n, \gamma)$.

[25] Sometimes wary beliefs are used in this context (see for instance Rey and Vergé, 2004; McAfee and Schwartz, 1994). With wary beliefs, a provider receiving contract (p_i, t_i) asks: given this contract, what are the cost minimizing contracts that I offers the other providers? As we have imposed little structure on our primitives, this question cannot be easily answered.

Competition between insurers makes sure that they choose n, γ to maximize consumers' valuation of insurance as given by (6) with $\sigma = C(n, \gamma)$:

$$V^i(n,\gamma) = \theta \int_{v(p(n,\gamma),\gamma)} (v-\gamma) f(v) dv - C(n,\gamma) - \theta\delta(p(n,\gamma),\gamma) \quad (23)$$

Using (18), we write this as

$$V^i(n,\gamma) = \theta \int_{v(p(n,\gamma),\gamma)} (v-c) f(v) dv - \Pi_P(\gamma, p_1(n,\gamma), \ldots, p_n(n,\gamma)) - \theta\delta(p(n,\gamma),\gamma) \quad (24)$$

In words, insurers maximize consumer welfare given by (i) the social value of treatment $(v-c)$ for all treatments that are done $(v > v(p(n,\gamma),\gamma))$; (ii) minus provider profits – as insurers earn zero profits, Π_P is paid by consumers and (iii) minus the dis-utility of risk aversion in case $\gamma > 0$. As insurers can offer $p = c = \gamma$ and contract all providers, we have $V^i(n, \gamma) \geq V^u$ in Eq. (5).

To ease notation, we use the following shorthands:

$$\bar{v}(n,\gamma) = v(p(n,\gamma),\gamma) \quad (25)$$

$$\bar{\delta}(n,\gamma) = \delta(p(n,\gamma),\gamma) \quad (26)$$

$$\bar{\Pi}_P(n,\gamma) = \Pi_P(\gamma, p_1(n,\gamma), \ldots, p_n(n,\gamma)) \quad (27)$$

$$\Delta\bar{\delta}(n,\gamma) = \bar{\delta}(n+1,\gamma) - \bar{\delta}(n,\gamma) \quad (28)$$

$$\Delta\bar{\Pi}_P(n,\gamma) = \bar{\Pi}_P(n+1,\gamma) - \bar{\Pi}_P(n,\gamma) \quad (29)$$

Now we can characterize the equilibrium n and γ set by insurers. In particular, as network size is an integer (the number of providers contracted), at the optimal network size n it must be the case that both a reduction and an increase in the number of providers contracted reduces welfare. Similarly, at the optimal co-payment γ, a change in γ must reduce welfare as well. Since $\gamma \geq 0$ (see Eq. (7)), a possible exception is the case where $\gamma = 0$ and a reduction in γ would raise welfare. Therefore, (32) allows for a strict inequality at $\gamma = 0$.

Proposition 3. *For given γ, the optimal network size n satisfies:*

$$\Delta\bar{\Pi}_P(n-1,\gamma) \leq \theta\left(\int_{\bar{v}(n,\gamma)}^{\bar{v}(n-1,\gamma)} (v-c) f(v) dv - \Delta\bar{\delta}(n-1,\gamma)\right) \quad (30)$$

$$\Delta\bar{\Pi}_P(n,\gamma) \geq \theta\left(\int_{\bar{v}(n+1,\gamma)}^{\bar{v}(n,\gamma)} (v-c) f(v) dv - \Delta\bar{\delta}(n,\gamma)\right) \quad (31)$$

For given n, the optimal co-payment γ is determined by

$$\frac{d\bar{\Pi}_P(n,\gamma)}{d\gamma} \geq \theta\left((c - \bar{v}(n,\gamma)) f(\bar{v}(n,\gamma))\bar{v}_\gamma(n,\gamma) - \bar{\delta}_\gamma(n,\gamma)\right) \quad (32)$$

where the inequality is strict at $\gamma = 0$ only.

Optimal network size is a trade-off between providers' profits and consumer utility. By increasing the network size, $p(n, \gamma) = p_n$ increases and patients are treated more often (Proposition 2). This increases the utility of treatment in case of under-treatment ($\bar{v} - c > 0$: the value of treatment exceeds the cost of treatment) and raises the dis-utility of risk aversion (in case $\gamma > 0$; recall that with $\gamma = 0$ there is no risk for the insured). The effect on provider profits can be both positive and negative.

Eq. (30) implies that moving from a network with $n-1$ to n providers increases V^i: consumer utility increases more than profits. Hence, this increase in the network size, increases welfare. But further increasing to $n+1$ reduces V^i: profits increase more than consumer utility in (31). Hence, the latter increase is not optimal and the optimal network size in n.

To get some more intuition, assume that the optimal $\gamma = 0$: as there is no co-payment for the insured, their risk aversion has no effect. That is, $\Delta\bar{\delta}(n, 0) = 0$ and the trade-off is between over/under-treatment and the effect of n on profits. Recall that the latter effect is non-monotone as $\Pi_P = 0$ both for $n = 1$ and for n high enough that $p_n = c$. Now Eq. (31) implies that under-treatment is possible (i.e. integral positive) in equilibrium if provider profits Π_P increase with n. Similarly, (30) implies that over-treatment $(\bar{v}(n-1,\gamma) < c)$ can happen if provider profits fall with n. By increasing n, the insurer induces inefficient treatment (where value is lower than treatment cost: $v - c < 0$) but as profits fall with n, the premium falls and consumers are better off. In this way, over-treatment can be an equilibrium outcome.

The optimal co-payment γ is also determined by the trade-off between over/under-treatment and consumer dis-utility on the one hand and provider profits on the other. If $\bar{v} < c$, an increase in γ reduces over-treatment which raises the value of insurance ex ante. In this case, $\gamma = 0$ can only be optimal if the costs $\theta\bar{\delta} + \bar{\Pi}_P$ increase fast with γ. If this is not the case, the optimal $\gamma > 0$ and (32) holds with equality.

With public contracts it is straightforward to implement first best $p = p^*$ and $\gamma = 0$. With private contracts this is not the case. First, the relation between p and n is determined by the insurer's cost side only (without reference to consumer valuation). Hence, generically speaking there is no n^* such that $p(n^*, 0) = p^*$. Even if such n^* would exist, Eqs. (30) and (31) imply that the effect of n on provider profits may induce $n \neq n^*$.

Hence, the model with private contracts explains why moral hazard in health care is still an unresolved issue. It cannot easily be resolved with a combination of demand and supply side measures. As a consequence, positive co-payments $\gamma > 0$ can be an equilibrium outcome. Although it is inefficient from a social point of view, supply side cost sharing cannot solve over-treatment in a model with private contracts.

Example 1 [continued]. Now we derive the network size and co-payment that maximizes V^i. With the parameter values introduced above, we find the following. With $n = 1$, we have $v(0, 0) = 0.60$, $C = 0.01$. Since $0.60 > 0.20 = c$, there is under-treatment and $\gamma = 0$. With $n = 2$, we get $p_2^2 = 0.08$, $v(0.08, 0) = 0.36$, $C = 0.02$. Since $0.36 > 0.2$, there is still under-treatment – hence $\gamma = 0$ with $n = 2$ – but less so than with $n = 1$. Consequently, total costs are higher with $n = 2$. The increase in treatment value equals

$$\int_{0.36}^{0.60} v f(v) dv = 0.08 \quad (33)$$

Hence, moving from $n = 1$ to $n = 2$ yields an increase in V^i equal to $0.08 - (0.02 - 0.01) > 0$. In this example, it is optimal for an insurer to contract both providers.

To illustrate the equilibrium, we give a simple formalization of the idea that people with higher income tend to buy insurance featuring a bigger network. To simplify notation, assume that the optimal $\gamma = 0$. Let $1/\mu$ denote the marginal utility of income, where people with higher income, have higher $\mu > 0$ and thus a lower marginal utility of income. Then we write Eq. (23) as

$$V^i = \theta \int_{v(p(n,0),0)} v f(v) dv - \frac{1}{\mu} C(n, 0) \quad (34)$$

Eq. (30) can be written as

$$\Delta\bar{\Pi}_P(n-1, 0) \leq \theta \int_{\bar{v}(n,0)}^{\bar{v}(n-1,0)} (v - \frac{c}{\mu}) f(v) dv \quad (35)$$

and higher μ leads to higher n as (35) is more easily satisfied for higher μ. Thus higher n can be optimal if income (μ) is higher. Note that in a cross section with agents with different μ's, different insurance contracts will be offered with varying network size. Each μ-type chooses the contract that is optimized for her; there are no incentive compatibility issues here.[26] In equilibrium, insurance contracts with a bigger network will feature higher fee-for-service. Hence, a patient is more likely to get treatment if her insurance contract features a bigger network.

5.2. Imperfect competition

In this subsection, we introduce two models with imperfect competition. The first shows that the result above that n and γ are chosen to maximize welfare can be extended to imperfect competition as well. The second one uses the example above to formalize that market power can lead insurers to choose a network that is more narrow than the one maximizing welfare.

First, assume that demand for insurer I_j takes a general form given by $Q^j(\hat{V}^j(n^j,\gamma^j)-\sigma^j,\hat{V}^{-j}(n^{-j},\gamma^{-j})-\sigma^{-j})$ where

$$\hat{V}^j(n,\gamma)=\theta\int_{v(p(n,\gamma),\gamma)}(v-\gamma)f(v)dv-\theta\delta(p(n,\gamma),\gamma). \tag{36}$$

That is, $\hat{V}^j=V^j+C^j$: the value for consumers of contract j not taking the cost/premium into account. Hence, the overall value for a consumer of contract j is given by $\hat{V}^j-\sigma^j$. Demand for j's contract is increasing in j's overall value ($\hat{V}^j-\sigma^j$ with a positive (own) partial derivative $Q^j_j>0$) and decreasing in value offered by a competitor $k\neq j$ ($\hat{V}^k-\sigma^k$ with a negative (cross) partial derivative $Q^j_k<0$). Insurer j then solves the following optimization problem.

$$\max_{\sigma^j,n^j,\gamma^j}(\sigma^j-C(n^j,\gamma^j))Q^j(\hat{V}^j(n^j,\gamma^j)-\sigma^j,\hat{V}^{-j}(n^{-j},\gamma^{-j})-\sigma^{-j}) \tag{37}$$

In equilibrium, each insurer chooses its premium, network size and co-payment optimally for given choices $n^{-j},\sigma^{-j},\gamma^{-j}$ by its competitors. As we show in the appendix, this implies the following result.

Lemma 1. *In equilibrium, each insurer I_j chooses n^j and γ^j to maximize welfare.*

The intuition is that in this set up, insurers maximize total welfare by setting n^j and γ^j and then use σ^j to capture the surplus from consumers. Hence, only σ^j is distorted in the sense that it exceeds marginal costs. As in the Bertrand case, insurers choose n and γ to maximize the "pie" and use σ to capture most of the rents generated. The intensity of competition on the insurance market determines how much σ exceeds marginal costs C.

Models like Hotelling competition have a demand structure that satisfies the assumptions above. Market power in itself does not need to cause a distortion in network size nor in demand side cost sharing.

To illustrate how market power can distort these dimensions, consider a monopolist insurer who faces consumers with different values for μ as in the example above. We assume that $\mu>0$ is distributed with density $g(\mu)$ and distribution function $G(\mu)$. Although consumers are differentiated, we do not allow the insurer to price discriminate. That is, the insurer offers one contract that can be bought by all consumers.

Writing utility as $\mu\hat{V}(n,\gamma)-\sigma$, we find that all types $\mu>\bar{\mu}$ buy insurance where $\bar{\mu}=\sigma/\hat{V}(n,\gamma)$. With perfect competition or a planner maximizing welfare, we have $\sigma=C(n,\gamma)$ and the marginal consumer is denoted by $\hat{\mu}=C(n,\gamma)/\hat{V}(n,\gamma)$. In this case, n, γ are chosen to maximize

$$\int_{\hat{\mu}}^{+\infty}(\mu\hat{V}(n,\gamma)-C(n,\gamma)g(\mu)d\mu. \tag{38}$$

The monopolist insurer chooses σ, n, γ to maximize its profits:

$$(1-G(\bar{\mu}))(\sigma-C(n,\gamma)) \tag{39}$$

The following result is proved in the appendix.

Lemma 2. *Compared to the planner, the monopoly insurer chooses a smaller network and higher out-of-pocket payment γ.*

The intuition for this result is the comparison of the valuation of the infra-marginal consumers under first best ($\hat{V}E(\mu|\mu>\hat{\mu})$, where E denotes the conditional expectation of μ) with the valuation of the marginal consumer only ($\hat{V}\bar{\mu}$). The planner takes into account that an increase in $\hat{V}$ benefits all consumers with high enough μ, while for the monopolist only the valuation of the marginal consumer matters as this consumer determines the level of the premium σ.

Since an increase in network size, increases utility (ignoring costs) such an increase in n has more weight for a planner than for a monopolist (i.e. the weight of the infra-marginal consumers exceeds the weight of the marginal consumer: $E(\mu|\mu>\bar{\mu})>\bar{\mu}$) while the cost effect is the same ($C_n(n,\gamma)$). Hence, a planner would implement a bigger network than a monopolist.

The opposite is true for γ: this decreases utility (ignoring costs) and hence it decreases utility more for the planner than for the monopolist for the same cost saving ($C_\gamma(n,\gamma)<0$). Consequently, the monopolist introduces more cost sharing than a planner would.

Hence, depending on how we model market power in the health insurance market, we find that network size and co-payment are chosen to maximize welfare as under Bertrand competition or that market power leads to a narrower network and higher demand side cost-sharing than a planner would like to implement. In either case, with market power the premium exceeds marginal costs thereby distorting insurance demand.

6. Policy implications

We have introduced a model where competing insurers use two instruments to control over-consumption in health care: demand-side and supply-side cost sharing. Whereas the former is publicly observable and contractible (insured need to know their co-payments) the latter is not. Neither consumers nor other providers dealing with an insurer know the details of the contract between the insurer and a provider, although this information is payoff relevant to them.

This has three implications. First, using "aggressive" capitation contracts (with a low fee-for-service) becomes expensive for an insurer as its network expands. Second, homogeneous providers earn strictly positive profits in a network with at least two providers and a fee-for-service below treatment cost. This gives them a motivation to defend the confidentiality clauses in their contracts with insurers. Third, insurers with a big network signal generous insurance to consumers: they pay a high fee-for-service making their physicians relatively willing to treat. This explains why insurers with bigger networks tend to have higher health care costs and utilization: patients are more likely to get treatment.

This section considers the implications of this model for two policies: AWP laws and initiatives to make prices more transparent.

[26] Analyzing adverse/advantageous selection in this model is beyond the scope of this paper. To illustrate why, assume that high risk types have a distribution function $F^h(v)$ and low risk types $F^l\geq F^h$. That is, conditional on falling ill, high risk types tend to draw higher v than low risk types (in the sense of first order stochastic dominance). Then single crossing may not be satisfied in this example. That is, at $n=1$, h-types may have a stronger preference for an increase in n than l-types, while for higher n the opposite is true. Indeed, for high n, the gain of a bigger network $F^i(\bar{v}(n-1,\gamma))-F^i(\bar{v}(n,\gamma))$ may be smaller for $i=h$ than for $i=l$. See Boone and Schottmüller (2017) for an analysis of health insurance when single crossing is not satisfied. As mentioned, Bardey and Rochet (2010) analyze optimal network size in a model with adverse selection.

We start by analyzing the effects of AWP laws. As a first observation, in reality contracts between providers and insurers are private (see Introduction). It is not clear how AWP laws deal with this. From a theory point of view, an insurer has to contract with any willing provider, but with private contracts it could just offer a loss making contract (low fee-for-service and low capitation) to providers it would rather not deal with. In practice, AWP laws "require managed care plans to explicitly state evaluation criteria and ensure "due process" for providers wishing to contract with the plan" (Hellinger, 1995, pp. 297). For the purpose of our model, let's assume that it is harder for an insurer to exclude a provider under AWP laws. That is, even though the insurer would rather not contract a provider, this provider can go to court and force a contract in a state with AWP laws.

Observe that with private contracts, provider profits can be strictly positive and hence an excluded provider has an incentive to join an insurer's network. This in contrast to public contract models in Section 3, where contracts can be chosen such that provider profits are zero. Hence, there is no strict incentive for a provider to join a network. Further with public contracts, even if more providers join a network, there is no effect on the health care costs of the network.

With private contracts, if an additional insurer joins the network, the fee-for-service tends to increase. Hence, if AWP laws lead to bigger networks, they will also increase health care consumption and costs. This is consistent with evidence cited in the Introduction. With Bertrand competition and homogeneous goods on the insurance market, insurers choose network size to maximize consumer value. AWP laws then unambiguously reduce welfare by forcing insurers to increase their network size. This conclusion does not necessarily follow with insurer market power as the network size may not be optimal in this case. To illustrate, we looked at an example where demand is such that insurer market power leads to networks that are too narrow from a social point of view. Then there can be a role for AWP laws to broaden these networks.

Discussions of policy initiatives to increase price transparency in health care provision usually focus on the bargaining effects and the risk of collusion (see e.g. Cutler and Dafny, 2011; Sinaiko and Rosenthal, 2011). There is agreement that patients should know the prices that they have to pay themselves; either through co-payments or by paying for uninsured treatments. But should there be transparency about the prices paid by insurers/managed care organizations to providers? Currently these prices are secret, what happens when they become public?

The bargaining effect from the literature can be illustrated as follows. Suppose that provider 1 has agreed to a low price with insurer A while charging another insurer B a high price. If these prices become public, insurer B will demand a low price as well, making provider 1 less likely to agree to such a low price in the first place. This suggests that price transparency raises prices. But this argument is not quite complete: suppose insurer B pays a low price to a provider 2; if prices become public, insurer B may be less willing to pay a high price to provider 1 as it fears that 2 will demand this high price as well. Hence, price transparency reduces the price insurer B pays to provider 1. In other words, the bargaining effect of price transparency is ambiguous.

The collusion argument goes as follows. Suppose providers try to form a cartel to charge insurers high prices. With secret contracts it is hard to detect a deviation from the cartel agreement making it hard to coordinate on a high price. With public prices, deviations are immediately detected and the cartel can sustain higher prices (Albaek et al., 1997).

Implicit in this analysis are two assumptions. First, prices are only transfers with no effects on health outcomes. Second, people are insured and therefore not interested in the prices that insurers pay to providers (conditional on the premium that they pay). As we have argued above, these assumptions are generally incorrect and therefore important effects of price transparency are overlooked.

First, as argued by the literature on supply-side cost sharing, prices are not neutral (see e.g. Ellis and McGuire, 1993; Ch et al., 2011). Higher fee-for-service leads to more treatments. Starting from a situation with under-treatment, such a price increase can be welfare enhancing. Our analysis above takes this effect into account.

Second, the prices paid by insurers to providers affect the insurance premium. The premium is clearly relevant for insured. But because of the previous effect, insured are interested in the prices paid to providers even ignoring the effect on the premium. Indeed, these prices affect the probability that the insured gets treatment when falling ill.

Taking these effects into account, what is the result of more price transparency in our framework? We argue that this depends on the degree of price transparency. If prices can become fully transparent to everyone, reforms to implement this transparency are welfare enhancing. Indeed, with public contracts it is possible to implement the first best. However, if prices only become partially transparent, we argue that these reforms reduce welfare. So when it comes to policies to increase price transparency the motto should be: do it well or not at all. This can be seen as follows.

Consider the case where prices become public to all providers, insurers and consumers. Then we are in the framework of Section 3 and the first best can be implemented. This clearly improves welfare compared to private contracts. However, making prices transparent to consumers is not going to be easy. We know that for treatments that patients actually use, they find it hard to understand what the price is (Rosenthal, 2014). Here we are considering a consumer who buys insurance – and therefore does not yet know what treatments he will need in the coming period – knowing all prices that his insurer will pay to the providers in the network. It is hard to envisage a policy that can increase transparency to such an extent.

The more likely effect will be that prices become transparent to providers and insurers but not to consumers buying insurance. Then the effect is that insurers can implement aggressive capitation contracts with a fee-for-service close to zero without leaving profits to insurers. The size of the network no longer signals the fee-for-service, as low fee-for-service can be implemented for any network size at no additional cost with public contracts. As the providers see each others' contracts, each expects to treat her share of the patients. They are willing to accept a contract with a capitation fee that just covers the expected loss per treatment. The effect of this degree of price transparency is that health care costs will fall. At first sight, the policy may then look successful. However, with very low fee-for-service fees there is under-provision of health care and hence the welfare consequences are not necessarily positive.

Summarizing, AWP laws tend to increase health care costs. If the insurance market is competitive, AWP laws tend to reduce welfare. Depending on the details of insurance demand, insurer market power can cause a bias towards narrow networks. In this case, AWP laws can raise welfare. If price transparency policies can make prices public to consumers, first best can be implemented and welfare increases. However, it seems more likely that prices become transparent to insurers and providers only. In that case, health care costs fall but so may welfare.

Appendix A. Proof of results

A.1 Proof of Proposition 1

Consider two offers $(p_i, t_i), (p_j, t_j) \in A_{\gamma,n}$ resp. with $p_j < p_i < c$ and other providers (in case $n \geq 3$) receive offers $p_k > p_i$. Hence $x_j = H(p_j,$

γ) and $x_i = H(p_i, \gamma) - H(p_j, \gamma)$. Then expected costs for the insurer are given by

$$C = x_i(p_i - \gamma) + \hat{x}_i(c - p_i) + x_j(p_j - \gamma) + \hat{x}_j(c - p_j) \tag{A.1}$$

One possible deviation is to offer both providers (p_i, t_i) – which is accepted as $(p_i, t_i) \in A_{\gamma,n}$. This leads to costs

$$\tilde{C} = (x_i + x_j)(p_i - \gamma) + 2\hat{x}_i(c - p_i) \tag{A.2}$$

There is no incentive to deviate if and only if $\tilde{C} - C \geq 0$. It is routine to verify that this can be written as

$$x_j(p_i - p_j) + \hat{x}_i(c - p^i) - \hat{x}_j(c - p_j) \geq 0 \tag{A.3}$$

Clearly, $(p_j, x_j(c - p_j)) \in A_{\gamma,n}$ with $\hat{x}_j = x_j = H(p_j, \gamma)$. Thus the following inequality needs to be satisfied:

$$-x_j(c - p_i) + \hat{x}_i(c - p_i) \geq 0 \tag{A.4}$$

Finally, let p_j approach p_i from below. Then this inequality implies that $\hat{x}_i \geq x_j$ which converges to $H(p_i, \gamma)$; that is, $\hat{x}_i \geq H(p_i, \gamma)$. □

A.2 Proof of proposition 2

As in the main text, we use the convention where $p_1 \leq p_2 \leq \ldots \leq p_n$. Define the function for provider profits $\Pi_P^*(n, p, \gamma)$ – which is different from (17) – as follows

$$\Pi_P^*(n, p, \gamma) = \min_{p_i \leq p} \sum_{i=1}^{n-1} H(p_i, \gamma)(c - p_{i+1}) \tag{A.5}$$

where p_n is optimally chosen such that $p_n = p$. Hence we find that

$$\frac{\partial \Pi_P^*(n, p, \gamma)}{\partial p} = -H(p_{n-1}, \gamma) < 0 \tag{A.6}$$

We can write

$$C(n, \gamma) = \min_{p_n} \Pi_P^*(n, p_n, \gamma) + H(p_n, \gamma)(c - \gamma) \tag{A.7}$$

The first order condition for an interior solution for p_n can be written as

$$-H(p_{n-1}, \gamma) + H_p(p_n, \gamma)(c - \gamma) = 0 \tag{A.8}$$

If the expression on the left hand side is positive at $p_n = 0$, then costs are minimized by choosing p_n as low as possible and we find $p_n = 0$ and consequently $p_1 = p_2 = \ldots = p_n = 0$. If the expression on the left hand side is negative at $p_n = c$, then costs are minimized by choosing $p_n \leq c$ as high as possible; that is $p_n = c$.

We need to establish the effects of γ, n on C and of n on p_n. First, consider the effect of γ on C. Using the envelope theorem, we have

$$\frac{\partial C(n, \gamma)}{\partial \gamma} = H_\gamma(p_n, \gamma)(c - \gamma) - H(p_n, \gamma) + \sum_{i=1}^{n-1} H_\gamma(p_i, \gamma)(c - p_{i+1}) < 0 \tag{A.9}$$

Second, to find the effect of n on C, start with $C(n, \gamma)$ and assume that prices $p_1 \leq \ldots \leq p_n < c$ minimize these costs. Moving from n back to $n - 1$, drop P_n's contract:

$$C(n-1, \gamma) \leq H(p_{n-1}, \gamma)(c - \gamma) + \sum_{i=1}^{n-2} H(p_i, \gamma)(c - p_{i+1}) < C(n, \gamma) \tag{A.10}$$

where the first inequality follows because $p_1, \ldots, p_{n-1}$ may not lead to lowest $C(n - 1, \gamma)$ and the second inequality follows from $H(p_{n-1}, \gamma) \leq H(p_n, \gamma)$, $\gamma < c$ and $H(p_n, \gamma)(c - p_n) > 0$.

Before, we derive the effect of n on p_n, we need the first order condition for an interior solution of p_i ($i \leq n - 1$) in (A.5):

$$H_p(p_i, \gamma)(c - p_{i+1}) - H(p_{i-1}, \gamma) = 0 \tag{A.11}$$

with $H(p_0, \gamma) = 0$ (as firms are indexed $i \geq 1$ and hence "provider" 0 – by convention – treats no patients). If the expression (on the left hand side) is positive, then p_i is chosen as low as possible: $p_i = p_{i-1}$. If it is negative, p_i is chosen as high as possible: $p_i = p_{i+1}$.

Next, we write

$$C(n+1, \gamma) = \min_{p_n \leq p_{n+1}} \Pi_P^*(n, p_n, \gamma) + H(p_n, \gamma)(c - p_{n+1}) + H(p_{n+1}, \gamma)(c - \gamma) \tag{A.12}$$

The derivatives of this expression with respect to p_n^{n+1}, p_{n+1}^{n+1} (price paid to P_n, P_{n+1} when the size of the network is $n + 1$) can be written as:

$$-H(p_n^{n+1}, \gamma) + H_p(p_n^{n+1}, \gamma)(c - p_{n+1}^{n+1}) \tag{A.13}$$

$$-H(p_n^{n+1}, \gamma) + H_p(p_{n+1}^{n+1}, \gamma)(c - \gamma) \tag{A.14}$$

The claim in the proposition is that $p_{n+1}^{n+1} \geq p_n^n$. Suppose – by contradiction – this is not the case: $p_{n+1}^{n+1} < p_n^n$, then we also have $p_n^{n+1} \leq p_{n+1}^{n+1} < p_n^n$. Hence evaluating (A.13) and (A.14) at $p_n^{n+1} = p_{n+1}^{n+1} = p_n^n$, it must be the case that both expressions are positive (i.e. evaluated at $p_n^{n+1} = p_{n+1}^{n+1} = p_n^n$, it is optimal to reduce p_n^{n+1}, p_{n+1}^{n+1}):

$$-H(p_n^n, \gamma) + H_p(p_n^n, \gamma)(c - p_n^n) > 0 \tag{A.15}$$

$$-H(p_n^n, \gamma) + H_p(p_n^n, \gamma)(c - \gamma) > 0 \tag{A.16}$$

Now consider two possibilities for the first order condition of p_n^n (A.8). First,

$$-H(p_{n-1}^n, \gamma) + H_p(p_n^n, \gamma)(c - \gamma) \leq 0 \tag{A.17}$$

Combining this with (A.16), leads to

$$-H(p_n^n, \gamma) + H(p_{n-1}^n, \gamma) > 0 \tag{A.18}$$

which is a contradiction because $H_p > 0$ and $p_n^n \geq p_{n-1}^n$. Second,

$$-H(p_{n-1}^n, \gamma) + H_p(p_n^n, \gamma)(c - \gamma) > 0 \tag{A.19}$$

Then p_n^n is chosen as low as possible: $p_n^n = p_{n-1}^n$. Let p_i^n denote the price for highest i such that

$$-H(p_{i-1}^n, \gamma) + H_p(p_i^n, \gamma)(c - p_{i+1}^n) \leq 0 \tag{A.20}$$

That is, all $j > i$ have a corner solution at the lower bound $p_j^n \geq p_{j-1}^n$: $p_n^n = \cdots = p_{i+1}^n = p_i^n$. Hence we can write (A.20) as

$$-H(p_{i-1}^n, \gamma) + H_p(p_n^n, \gamma)(c - p_n^n) \leq 0 \tag{A.21}$$

Combining this with (A.15) leads to a contradiction because $p_{i-1}^n \leq p_n^n$ implies $H(p_n^n, \gamma) \geq H(p_{i-1}^n, \gamma)$.

Finally, consider the effect of γ on p_n. If p_n is a corner solution, then a small change in γ has no effect on p_n. If p_n is characterized by first order condition (A.8), then the second order condition (for a minimum) implies $H_{pp}(p_n, \gamma)(c - \gamma) > 0$. Hence

$$\text{sign} \left(\frac{dp_n}{d\gamma}\right) = \text{sign} \left(H_\gamma(p_{n-1}, \gamma) + H_p(p_n, \gamma) - H_{p\gamma}(p_n, \gamma)(c - \gamma)\right) \tag{A.22}$$

Since $H_\gamma < 0$ and $H_p > 0$, we cannot sign this expression in general. Moreover, we did not make an assumption on $H_{p\gamma}$. Hence, the effect of γ on p_n ambiguous. □

A.3 Proof of Lemma 1

To simplify the exposition, we will take the derivative with respect to network size n. This is a slight abuse of notation as n is actually an integer. However, it does make clear the structure of the problem and why we get efficiency in this case.

The first order conditions of the problem (37) can be written as

$$Q^j(.) - Q^j_j(.)(\sigma^j - C(n^j, \gamma^j) = 0 \tag{A.23}$$

$$-C_n(n^j, \gamma^j)Q^j(.) + (\sigma^j - C(n^j, \gamma^j))Q^j_j(.)\hat{V}^j_n(n^j, \gamma^j) = 0 \tag{A.24}$$

$$-C_\gamma(n^j, \gamma^j)Q^j(.) + (\sigma^j - C(n^j, \gamma^j))Q^j_j(.)\hat{V}^j_\gamma(n^j, \gamma^j) = 0 \tag{A.25}$$

Combining the first two equations, allows us to write

$$\hat{V}^j_n(n^j, \gamma^j) - C_n(n^j, \gamma^j) = 0 \tag{A.26}$$

In words, I_j chooses network size n^j to maximize welfare. Combining the first and last equation, we get

$$\hat{V}^j_\gamma(n^j, \gamma^j) - C_\gamma(n^j, \gamma^j) = 0 \tag{A.27}$$

Demand side cost sharing is chosen to maximize welfare as well. □

A.4 Proof of Lemma 2

As in the previous proof, we abuse notation by differentiating with respect to n. The first order condition of (38) with respect to n can be written as

$$\hat{V}_n(n, \gamma)\int_{\hat{\mu}}^{+\infty} \mu g(\mu)d\mu - (1 - G(\hat{\mu}))C_n(n, \gamma) = 0 \tag{A.28}$$

which we write as

$$\hat{V}_n(n, \gamma)E(\mu|\mu > \hat{\mu}) - C_n(n, \gamma) = 0 \tag{A.29}$$

where

$$E(\mu|\mu > \hat{\mu}) = \frac{\int_{\hat{\mu}}^{+\infty} \mu g(\mu)d\mu}{1 - G(\hat{\mu})} \tag{A.30}$$

Similarly, the first order condition for γ can be written as

$$\hat{V}_\gamma(n, \gamma)E(\mu|\mu > \hat{\mu}) - C_\gamma(n, \gamma) = 0 \tag{A.31}$$

The monopolist maximizes Eq. (39) with $\bar{\mu} = \sigma/\hat{V}(n, \gamma)$. The first order condition with respect to σ can be written as

$$-g(\bar{\mu})\frac{1}{\hat{V}(n, \gamma)}(\sigma - C(n, \gamma)) + (1 - G(\bar{\mu})) = 0 \tag{A.32}$$

Combining this with the first order condition of n, we can write the latter as

$$\bar{\mu}\hat{V}_n(n, \gamma) - C_n(n, \gamma) = 0 \tag{A.33}$$

Similarly, the first order condition for γ can be written as

$$\bar{\mu}\hat{V}_\gamma(n, \gamma) - C_\gamma(n, \gamma) = 0 \tag{A.34}$$

for given choice of $\sigma > C$, the planner would take into account that infra-marginal consumers value an increase in n more than the marginal consumer: $E(\mu|\mu > \bar{\mu}) > \bar{\mu}$ and hence would choose a bigger network than the monopolist.

In contrast, $\hat{V}_\gamma < 0$ (there is a cost-saving $C_\gamma < 0$) and the planner gives a bigger weight to the welfare loss than the monopolist. Hence, the planner would like to reduce γ compared to the monopolist. □

References

Albaek, S., Mollgaard, P., Overgaard, P.B., 1997. Government-assisted oligopoly coordination? A concrete case. J. Ind. Econ. 45 (December (4)), 429–443.
Aron-Dine, A., Einav, L., Finkelstein, A., 2013. The rand health insurance experiment, three decades later. J. Econ. Perspect. 27 (1), 197–222.
Arrow, K., 1963. Uncertainty and the welfare economics of medical care. Am. Econ. Rev. 53 (5), 941–973.
Bardey, D., Lesur, R., 2006. Optimal regulation of health system with induced demand and ex post moral hazard. Ann. d'Écon. Stat. 83/84, 279–293.
Bardey, D., Rochet, J.-C., 2010. Competition among health plans: a two-sided market approach. J. Econ. Manag. Strategy 19 (2), 435–451, ISSN 1530-9134.
Barros, P.P., Martinez-Giralt, X., 2008. Selecting health care providers: "any willing provider" vs. negotiation. Eur. J. Polit. Econ. 24 (June (2)), 402–414, ISSN 0176-2680.
Bernheim, B.D., Whinston, M.D., 1998. Exclusive dealing. J. Polit. Econ. 106 (1), 64–103.
Boone, J., Douven, R., 2014. Provider Competition and Over-Utilization in Health Care. Technical Report, CEPR DP No. DP10177.
Boone, J., Schottmüller, C., 2017. Health insurance without single crossing: why healthy people have high coverage. Econ. J. 127 (599), 84–105.
Boone, J., Schottmüller, C., 2019. Do health insurers contract the best providers? Provider networks, quality, and costs. Int. Econ. Rev. 60 (August (3)), http://dx.doi.org/10.1111/iere.12383.
Capps, C., Dranove, D., Satterthwaite, M., 2003. Competition and market power in option demand markets. RAND J. Econ. 34 (Winter (4)), 737–763.
Chandra, A., Cutler, D., Song, Z., 2011. Chapter six – who ordered that? the economics of treatment choices in medical care. In: Mcguire, T.G., Pauly, M.V., Barros, P.P. (Eds.), Handbook of Health Economics, volume 2 of Handbook of Health Economics. Elsevier, pp. 397–432.
Chernew, M., DeCicca, P., Town, R., 2008. Managed care and medical expenditures of medicare beneficiaries. J. Health Econ. 27 (6), 1451–1461, ISSN 0167-6296.
Chernew, M.E., Newhouse, J.P., 2011. Chapter one – health care spending growth. In: Mcguire, T.G., Pauly, M.V., Barros, P.P. (Eds.), Handbook of Health Economics, volume 2 of Handbook of Health Economics. Elsevier, pp. 1–43.
Christianson, J.B., Conrad, D., 2011. Chapter 26 – provider payment and incentives. In: Glied, S., Smith, P. (Eds.), Oxford Handbook of Health Economics. Oxford University Press, pp. 624–648.
Coe, E., Leprai, C., Oatman, J., Ogden, J., 2013. Hospital Networks: Configurations on the Exchanges and their Impact on Premiums. Technical Report. McKinsey Center for U.S. Health System Reform.
Cutler, D., Reber, S., 1998. Paying for health insurance: the tradeoff between competition and adverse selection. Q. J. Econ. 113 (2), 433–466.
Cutler, D., Dafny, L., 2011. Designing transparency systems for medical care prices. New Engl. J. Med. 364 (10), 894–895, PMID: 21388307.
Cutler, D.M., McClellan, M., Newhouse, J.P., 2000. How does managed care do it? RAND J. Econ. 31 (3), 526–548, ISSN 07416261.
Cutler, D.M., 2004. Your Money or Your Life: Strong Medicine for America's Healthcare System. Oxford University Press.
De La Merced, M.J., 2014. Oscar, a New Health Insurer, Raises $30 Million (accessed 05.05.14) http://dealbook.nytimes.com/2014/01/07/oscar-a-new-health-insurer-raises-30-million/?_php=true&_type=blogs&_r=0.
Dranove, D., 2000. The Economic Evolution of American Health Care: From Marcus Welby to Managed Care. Princeton University Press.
Dranove, D., Shanley, M., White, W., 1993. Price and concentration in hospital markets: the switch from patients-driven to payer-driven competition. J. Law Econ. 36 (April (1)), 179–204.
Ellis, R.P., McGuire, T.G., 1993. Supply-side and demand-side cost sharing in health care. J. Econ. Perspect. Fall 7 (4), 135–151.
Gal-Or, E., 1997. Exclusionary equilibria in health-care markets. J. Econ. Manag. Strategy 6 (1), 5–43.
Gaynor, M., Ho, K., Town, R.J., 2015. The industrial organization of health-care markets. J. Econ. Lit. 53 (June (2)), 235–284.
Haas-Wilson, D., 2003. Managed Care and Monopoly Power: The Antitrust Challenge. Harvard University Press.
Hart, O., Tirole, J., 1990. Vertical integration and market foreclosure. Brook. Papers Econ. Act, 205–276.
Hellinger, F.J., 1995. Any-willing-provider and freedom-of-choice laws: an economic assessment. Health Aff. 14 (4), 297–302.
Ho, K., 2009. Insurer-provider networks in the medical care market. Am. Econ. Rev. 99 (1), 393–430.
Ho, K., Pakes, A., 2013. Hospital Choices, Hospital Prices and Financial Incentives to Physicians. Working Paper 19333. National Bureau of Economic Research, August.
Kirk, P., 2014. Health Insurers Encourage Physicians to Help Patients Use Cost and Quality Data to Select Providers, Including Medical Laboratories (accessed 05.05.14) http://www.darkdaily.com/health-insurers-encourage-physicians-to-help-patients-use-cost-and-quality-data-to-select-providers-including-medical-laboratories.
Klick, J., Wright, J.D., 2014. The effect of any willing provider and freedom of choice laws on prescription drug expenditures. Research Paper 12-39. U of Penn, Inst for Law & Econ, February.
Kockesen, L., 2007. Unobservable contracts as precommitments. Econ. Theory 31 (3), 539–552.
Laffont, J.J., Tirole, J., 1993. A Theory of Incentives in Procurement and Regulation. MIT Press.

Lesser, C., Ginsburg, P.B., Devers, K., 2003. The End of an Era: What Became of the "Managed Care Revolution" in 2001? Health Services Research.
Liu, D., 2013. An Excellent Primary Care Doctor is Your Trusted Health Care Advisor (accessed 05.05.14) http://www.kevinmd.com/blog/2013/01/excellent-primary-care-doctor-trusted-health-care-advisor.html.
Ma, C.-T.A., McGuire, T.G., 1997. Optimal health insurance and provider payment. Am. Econ. Rev. 87 (September (4)), 685–704.
Ma, C.-T.A., McGuire, T.G., 2002. Network incentives in managed health care. J. Econ. Manag. Strategy Spring 11 (1), 1–35.
Ma, C.A., Riordan, M.H., 2002. Health insurance, moral hazard, and managed care. J. Econ. Manag. Strategy 11 (Spring (1)), 81–107.
McAfee, P., Schwartz, M., 1994. Opportunism in multilateral vertical contracting: nondiscrimination, exclusivity, and uniformity. Am. Econ. Rev. 84, 210–230.
McGuire, T.G., 2011. Chapter five - demand for health insurance. In: Mcguire, T.G., Pauly, M.V., Barros, P.P. (Eds.), Handbook of Health Economics, volume 2 of Handbook of Health Economics. Elsevier, pp. 317–396.
Muir, M.A., Alessi, S.A., King, J.S., 2013. Clarifying Costs: Can Increased Price Transparency Reduce Healthcare Spending? Technical Report 38, UC Hastings Research Paper, February.
Pear, R., 2014. To prevent surprise bills, new health law rules could widen insurer networks. New York Times (July).
Posnett, J., 1999. Is bigger better? concentration in the provision of secondary care. BMJ 319 (16), 1063–1065, October 1999.
Rey, P., Vergé, T., 2004. Bilateral control with vertical contracts. RAND J. Econ. 35 (Winter (4)), 728–746.
Rosenthal, E., 2014. Patients' Costs Skyrocket; Specialists' Incomes Soar (accessed 05.05.14) http://www.nytimes.com/2014/01/19/health/patients-costs-skyrocket-specialists-incomes-soar.html.
Scott, G., 2011. Insurers Steer Patients Away From Pricey ERS (accessed 05.05.14) http://www.crainsnewyork.com/article/20110815/FREE/110819936/insurers-steer-patients-away-from-pricey-ers.
Segal, I., 1999. Contracting with externalities. Q. J. Econ. 114, 337–388.
Segal, I., Whinston, M.D., 2003. Robust predictions for bilateral contracting with externalities. Econometrica 71, 757–791.
Sinaiko, A.D., Rosenthal, M.B., 2011. Increased price transparency in health care – challenges and potential effects. New Eng. J. Med. 364 (10), 891–894, PMID: 21388306.
Terhune, C., 2013. Insurers limiting doctors, hospitals in health insurance market. Los Angeles Times (September).
Zwanziger, J., Melnick, G.A., 1996. Can managed care plans control health care costs? Health Aff. 15 (2), 185–199.
Zwanziger, J., Melnick, G.A., Bamezai, A., 1994. Costs and price competition in California hospitals, 1980–1990. Health Aff. 13 (4), 118–126.
Zwanziger, J., Melnick, G.A., 1988. The effects of hospital competition and the medicare pps program on hospital cost behavior in california. J. Health Econ. 7 (4), 301–320.