RAND Journal of Economics
Vol. 45, No. 4, Winter 2014
pp. 792–815

Competition leverage: how the demand side affects optimal risk adjustment

Michiel Bijlsma*

Jan Boone**

and

Gijsbert Zwart*

We study optimal risk adjustment in imperfectly competitive health insurance markets when high-risk consumers are less likely to switch insurer than low-risk consumers. Insurers then have an incentive to select even if risk adjustment perfectly corrects for cost differences. To achieve first best, risk adjustment should overcompensate insurers for serving high-risk agents. Second, we identify a trade-off between efficiency and consumer welfare. Reducing the difference in risk adjustment subsidies increases consumer welfare by leveraging competition from the elastic low-risk market to the less elastic high-risk market. Third, mandatory pooling can increase consumer surplus further, at the cost of efficiency.

1 Introduction

■ Buyers of health insurance differ greatly in their expected health care costs. Even if some of their health characteristics are observable, community rating requirements prohibit health insurers from exploiting this information. By stipulating that insurance contracts on offer should be available to all consumers at the same price, such requirements prevent insurers from practicing explicit price discrimination. Insurance companies can, however, engage in second-degree price discrimination and offer contracts with under insurance to cherry pick low-risk consumers. In the presence of such risk selection, consumers with high expected costs get efficient insurance but pay high premiums. Moreover, such separation of high-cost and low-cost consumers through

* TILEC, Tilburg University, and Netherlands Bureau for Economic Policy Analysis, CPB; m.j.bijlsma@cpb.nl, g.t.j.zwart@cpb.nl.
** CentER, Tilburg University and TILEC, CPB, CEPR; j.boone@uvt.nl.
We thank David Martimort, the editor, and the referees, as well as Rudy Douven and seminar audiences at the University of Groningen, January 2011, and at the Dutch Healthcare Regulator, NZa, February 2011, for comments and suggestions. Boone gratefully acknowledges financial support from the Netherlands Organisation for Scientific Research (NWO) through a Vici grant (no. 453.07.003).

contracts with distorted coverage leads to inefficiency in health insurance markets, as explained by Rothschild and Stiglitz (1976).[1]

As a policy response to risk selection, and the efficiency and equity problems it creates, many countries have introduced risk adjustment schemes. Such schemes consist of transfers to insurers designed to reduce selection incentives by eliminating differences in expected costs among consumers.[2] Designing those schemes involves predicting the expected health care costs of an individual based on observed characteristics, such as age, gender, and previous diagnoses and drug prescriptions.[3] Risk adjustment schemes involve significant amounts of money. To illustrate, the Netherlands currently has an elaborate nationwide risk adjustment system with a 2011 budget equal to 18 billion euro, which amounts to approximately 3% of GDP (Gross Domestic Product).

Risk adjustment payments that are based on differences in consumers' expected health care costs are often argued to help in attaining two separate policy goals (Sass, 1995; Van de Ven and Ellis, 2000; Olson, 2011): (i) increasing efficiency by reducing incentives for risk selection and (ii) increasing solidarity by cross-subsidizing from low-risk to high-risk consumers. The claim is that once health care costs for an individual can be perfectly predicted and risk adjustment fully corrects for any predictable cost differences (we call this *perfect risk adjustment*), insurance companies have no incentives to engage in risk selection by distorting contract terms. Second, risk adjustment is claimed to enhance equity, as there is a higher benefit for those who have high health risks, and hence insurance premiums for those types are reduced. Consequently, in the words of Olson (2011): "one's contribution to the financing of health care is divorced from one's needs for health care."[4,5] Glazer and McGuire (2000), Glazer and McGuire (2002), and Jack (2006) present theoretical models that bear out these claims.

We show that defining risk adjustment payments based on expected costs only is unlikely to optimize efficiency or equity in reality, if insurers have market power and expected health costs correlate with the propensity to switch insurer. In practice, consumers with high expected costs are often less likely to switch insurers.[6] For example, in a US study, Strombom, Buchmueller, and Feldstein (2002) find that young and healthy employees are four times more price elastic than the oldest employees. Schut, Gress, and Wasem (2004) find a similar ratio of elasticities in a comparison between German nonpensioners and pensioners. Other studies documenting that old or sick insured are less likely to switch insurer include Beaulieu (2002), Buchmueller (2006), Nuscheler and Knaus (2005), Royalty and Solomon (1999), and Van Vliet (2006). This connection between health costs and price elasticity is overlooked in the traditional literature on risk adjustment. The main contribution of our article is to model explicitly this relation between expected costs and tendency to switch insurer, and to analyze the implications for the design of the risk adjustment scheme.

We consider a model of two horizontally differentiated insurers, competing for consumers that are heterogeneous in health risks (along the lines of Stole, 1995). In keeping with the empirical evidence, we model residual demand elasticity of high-risk consumers to be lower than that of low–risk ones. We assume that insurers can observe consumer risk types but cannot use this information as a consequence of a community rating requirement. They can, however, distort

[1] For an empirical verification of these predictions, see Buchmueller and DiNardo (2002). Cutler and Reber (1998) provide an estimate of the costs of such inefficiencies.

[2] Depending on the way that risk adjustment is organized, the transfers to insurers are paid by the government or another sponsor of the scheme, like an employer.

[3] Van de Ven and Ellis (2000) give an overview of variables used in risk adjustment.

[4] Schokkaert, Dhaene, and Van de Voorde (1998), building on the literature on fair compensation (e.g., Bossert and Fleurbaey, 1996), discuss risk adjustment as a means to implement equitable contributions to insurance from individuals, that is, contributions not depending on characteristics for which the individual does not bear responsibility.

[5] The argument of solidarity is also often used to motivate community rating, although it is not clear that community rating is the right instrument for this goal (Van de Ven and Schut, 2011).

[6] This is related to adverse retention as discussed by Cutler, Lincoln, and Zeckhauser (2010). A possible explanation for this relation between expected costs and switching behavior is that those already undergoing treatment are more strongly locked into their relationships with both contracted physicians and insurers.

coverage of the policies they offer to induce consumers to self-select into contracts intended for them.[7]

The planner similarly has complete information on consumer types.[8] It uses that information to set up a risk adjustment scheme consisting of taxes and subsidies to insurers that depends on the consumers they attract. These risk adjustment payments will, in turn, affect the pricing and coverage of the contracts that insurers will offer in equilibrium. We thus study a two-tiered structure in which, first, the planner sets up the risk adjustment scheme, and second, insurers compete in contracts for consumers, taking this regulatory background as given.

We ask how the planner should set risk adjustment payments if it maximizes a weighted welfare function, where we allow for different weights on high-risk and low-risk consumers, as well as on consumers versus insurers, to reflect different preferences for redistribution on the part of the planner.

A change in the risk adjustment payments offered by the planner will induce a change in the equilibrium level of coverage offered by the insurers. We first analyze what level of coverage the planner will want to attain. If the planner optimizes unweighted total welfare, it will strive to eliminate any distortions in coverage and implement first-best insurance. However, if the planner places higher weight on consumer surplus than on insurer surplus, an opposing effect kicks in, and coverage for low types will be below first-best levels. The intuition is that increasing the risk adjustment subsidy on the low types above their first-best levels induces insurers to compete more vigorously for these low-risk, but more elastic types. More vigorous competition for low types reduces their insurance premium. However, high-risk types also benefit from intensified competition in the low-risk market via the incentive compatibility (IC) constraint that connects the two groups' premiums. Therefore, biasing the risk adjustment transfers against the high types leverages competition from the low-risk market, where consumers are likely to switch insurers, to the high-risk market, where the tendency to switch insurers is lower. Starting from first best, these price effects are first order and dominate the second-order welfare loss from the decrease in coverage that results from the insurers' selection incentives.

Next, we explore how the planner should set risk adjustment payments in order to achieve the desired equilibrium level of coverage. We consider first the case in which the planner wants to implement first-best coverage. According to the existing literature, this is achieved by fully correcting for differences in consumers' expected costs. Instead, we show that insurers still have an incentive to select in that case: risk adjustment also needs to compensate insurers for the difference in markups (the ratio of price and elasticity) between the less elastic high types and the more elastic low types. To get the first-best outcome, the planner should therefore overcompensate for the high-risk types: the difference in risk adjustment transfers between the high and low type should exceed the difference in expected costs and include the difference in mark-ups. Assuming elasticities differ a factor of four (see Strombom, Buchmueller, and Feldstein, 2002; Schut, Gress, and Wasem, 2004), this correction to traditional risk adjustment is of the same order of magnitude as the markup itself and thus significant in less than fully competitive insurance markets.

If the planner puts more weight on consumers than on insurers, it wants to lower risk adjustment contributions. Low types, and by competition leverage also the high types, then benefit from the resulting, more intense, competition. This margin effect is first order when consumer surplus gets higher weight than producer surplus, and it outweighs the second-order effect of the reduced coverage for low types compared to first-best levels. To illustrate the power of competition leverage, we derive conditions under which maximizing the utility of the high type implies that the risk adjustment scheme taxes the high-risk type and subsidizes the low type, the opposite direction of traditional risk adjustment.

[7] In our analysis, insurers can raise prices sufficiently to also offer a profitable high-coverage contract to high risks. If this is not possible, quality of low-risk contracts will be further degraded in an effort to drive away high risks.

[8] In Section 7, we relax the assumption that the planner and insurers perfectly observe risk types, allowing instead for imperfect signals of consumer types.

Finally, we analyze whether the planner can do better by enforcing mandatory pooling and forbidding health insurers to offer menus of contracts. If people differ only in expected costs, but not in switching elasticities, optimal risk adjustment implies perfect equalization of the various types' expected costs. Each cost type then gets efficient insurance and (in the Rothschild-Stiglitz model) low and high types are offered the same contracts. Hence, enforcing mandatory pooling creates no added benefit. However, in our model with different switching behavior, mandatory pooling has a benefit. In particular, mandatory pooling tightens the constraint relating high to low type utility. As a result, a given level of competition leverage comes at lower distortion costs. Hence, consumers are better off under mandatory pooling than if firms are allowed to engage in second-degree price discrimination. In practice, this welfare gain for consumers needs to be weighed against reduced variety in contracts, which may reduce welfare if consumers have different tastes for risks.

Our article is related to the health economics literature on health insurance and risk adjustment and the industrial organization (IO) literature on oligopoly with price discrimination. We discuss each in turn. The empirical literature on risk adjustment tries to find sets of variables that explain the individual's next period health care costs, so that the remaining individual variance in costs is minimized. It focuses on variables that are easily available, such as age, gender, or zip code, so as to ensure that the estimated model can be used in practice. Theoretical articles, like Glazer and McGuire (2000), Glazer and McGuire (2002), and Jack (2006) analyze how imperfect signals about consumer types should be mapped into risk adjustment transfers. In these articles, the absence of noise implies that perfect compensation for expected costs leads to efficient market outcomes. Our article differs from these articles by explicitly modelling the relation between risk type and elasticity to switch insurers, in a context of imperfect competition among insurers. Recent evidence that market power is important in health insurance markets includes Dafny (2010). Related theoretical models of insurer competition are Olivella and Vera-Hernández (2007) and Jack (2006). We focus on the difference in elasticity among the various consumer types and show that, even with perfectly informative signals, perfect risk adjustment is inefficient.

Related to our work, Jack (2001) also considers a model where risk adjustment can have adverse consequences for consumers. In Jack (2001), competing insurers offer a pooling contract to consumers of two types, as in our analysis of mandatory pooling. Insurers screen consumers by reducing quality. In addition, insurers can exert effort to reduce the cost in providing insurance to high-cost consumers. Risk adjustment, which in Jack (2001) takes the form of a compensation of *ex post* realized costs, now has two opposing effects. On the one hand, it covers part of the additional costs incurred for high risks. On the other hand, it lowers incentives to reduce costs. Jack (2001) shows that the latter, moral hazard effect can outweigh the cost equalization effect and then more risk adjustment decreases quality. In contrast to Jack (2001), we study type-contingent *ex ante* risk adjustment, which does not reduce incentives for cost reduction. In our work, positive risk adjustment always reduces screening. However, as a result of the difference in elasticities, this reduction in screening goes together with a decrease in competition intensity, and hence an increase in margins.

In the industrial economics literature a number of articles study price discrimination in an oligopoly setting. Examples include Stole (1995), Armstrong and Vickers (2001), and Schmidt-Mohr and Villas-Boas (1999) (see Stole, 2007, for an overview). These articles study equilibria where competing principals (the sellers) offer contracts to consumers in an asymmetric information setting. In contrast, in this article, we assume that insurers can observe consumer types but are prevented from using that information by a community rating requirement.[9] The novelty in our work is that we include an additional hierarchical layer, the planner, who has equal information on consumer types and can use that to implement a regulatory scheme taxing and subsidizing insurers based upon the types of the consumers they attract. We then explore how the planner

[9] In Section 7, we extend the analysis to the case when observation is imperfect.

can achieve its goals by choosing this scheme to optimally influence the insurers' contracting equilibrium.

In this latter respect, our article touches on the literature on two-tiered hierarchies, where intermediaries (the insurers) contract with agents (the consumers), on behalf of a principal (the planner). An example in this line of research related to our work is DeMarzo, Fishman, and Hagerty (2005), who consider a principal's choice of an antifraud enforcement mechanism, and show how this choice affects competition for customers among security dealers. DeMarzo, Fishman, and Hagerty (2005) show that a self-regulatory organization may use this mechanism to mute competition among their members. Although that article looks at a model with moral hazard and costly state verification, Faure-Grimaud, Laffont, and Martimort (1999) consider a hierarchical model where the intermediary (an auditor) offers a menu of contracts to a manager who can be of two types. The auditor himself is hired by the principal (the shareholders). The contract between shareholder and auditor will again influence the menu of contracts offered by the auditor, as in this article. A difference with Faure-Grimaud, Laffont, and Martimort (1999) is that our article looks at competition among intermediaries, as do DeMarzo, Fishman, and Hagerty (2005) in their moral hazard context. Moreover, unlike these articles, we consider a situation where information on consumer types is symmetric; the nontrivial contracting structure is a consequence of legal restrictions on price discrimination.

The setup of our article is as follows. We first provide some descriptive background on insurer selection and the risk adjustment schemes that are used in practice by sketching the system used in the Netherlands. Next, we introduce the model in Section 3. Section 4 explores what level of coverage optimizes the planner's objective function, and Section 5 then finds how risk adjustment payments should be set to achieve that optimum. Section 6 shows that mandatory pooling may increase the planner's welfare. Section 7 shows that our results are robust to noisy signals. We conclude with a discussion of policy implications.

2 Risk adjustment in practice

■ Risk adjustment is used in many countries in different degrees of sophistication. Countries that use risk adjustment include Belgium, Colombia, Germany, Israel, the Netherlands, Switzerland, and the United States. Van de Ven and Ellis (2000), Ellis (2008), and Armstrong et al. (2010) discuss the different ways risk adjustment is used in these countries. In the United States, the health insurance exchanges that are currently being established under the Patient Protection and Affordable Care Act will include a risk adjustment scheme (HHS, 2012). For concreteness, we describe the health insurance market and its current risk adjustment system in the Netherlands. Dutch risk adjustment is viewed as one of the most sophisticated systems in the world (Stam, 2007).

The Netherlands has mandatory health insurance for basic insurance. In addition to this, there is voluntary supplementary insurance that covers treatments not covered by basic insurance, such as dental care and physiotherapy. A good overview is provided by van de Ven and Schut (2008).

Dutch insurers have to accept any applicant for the basic package at a community-rated premium. The Dutch Health Insurance Act prescribes the functions of care that need to be covered by the basic package, and it sets a minimum deductible. Such regulations do not apply to supplementary insurance: there is no mandatory acceptance, and insurers can freely determine the care insured under these supplementary policies. In the Netherlands, more than 90% of the population buy supplementary insurance, typically from the same insurer that they bought the basic package from.

As we know from Van de Ven and Ellis (2000) (and as we will see below), community rating creates incentives for risk selection among insurers. Although the insurers have to insure all care in the basic package, they still have ample opportunity for such selection; see van de Ven and Schut (2008). As the basic package does not prescribe the providers where patients can be treated, insurers can use selective contracting to determine their provider network. Insurance with a low premium and a narrow network is likely to attract relatively more low-risk types compared to more expensive insurance that covers all providers in the Netherlands.[10,11] Also, insurers can offer contracts with a higher deductible than the minimum specified by law to select for less costly types. As people tend to buy basic and supplementary insurance from the same insurer, the generosity of supplementary insurance (which is subject to less regulation) can also be used to select risk types for basic insurance.

To limit such risk selection, the Dutch government uses risk adjustment, making payments out of a risk equalization fund, funded by all insured. In general, these payments can be determined in two ways: *Ex post* (retrospective) and *ex ante* (prospective). *Ex post* contributions from the government to an insurer are based on the insurer's realized cost: the government then reinsures the insurers for part of their realized expenses. Clearly, by reinsuring insurers completely, all selection incentives disappear. However, all incentives to reduce costs through, say, bargaining with providers about prices and decreasing overconsumption of care disappear as well.[12] For this reason, the Dutch government has decided to abolish almost all of the *ex post* risk adjustment.

Conversely, the extent of *ex ante* risk adjustment has grown strongly over recent years. *Ex ante* risk adjustment uses an insured agent's observable characteristics at the start of the period to predict her costs during the period. It therefore avoids the adverse effects on insurer incentives for cost containment that are associated with *ex post* adjustment. The goal of *ex ante* risk adjustment is to choose those explanatory variables that best predict expected costs, minimizing remaining individual variance among the insured. Compensating expected costs takes away insurer incentives to select through distortions in contracts that are offered, and in that way, *ex ante* risk adjustment increases efficiency of the market equilibrium. Initially, Dutch *ex ante* risk adjustment consisted only of corrections for age, gender, and some socio economic indicators. In recent years, factors correcting for chronic illnesses have been added, such as previous years' diagnoses and previous use of pharmaceuticals, and new explanatory variables continue to be explored.[13] The aim of the system has been to make sure that insurers cannot make predictable profits or losses on identifiable subgroups of consumers (see Van de Ven and Schut, 2008).[14]

The model below can be applied in any regulatory environment where community rating causes insurers to cherry pick insured using second-degree price discrimination. Risk adjustment then affects selection incentives and hence, welfare. If community rating is not imposed and insurers are allowed to risk rate, there is no incentive compatibility constraint. The outcome is then always efficient and risk adjustment only affects prices. Then there is no competition leverage. See Bijlsma, Boone, and Zwart (2013) for an analysis of the conditions under which community rating is optimal for the planner to impose.

3 The model

■ We consider a two-tier model of a planner setting a risk adjustment scheme, and two health insurers competing for consumers that can be of two types, high or low cost. First, the planner determines the risk adjustment scheme. Next, taking this scheme as given, the imperfectly

[10] As an example, the Dutch insurance brand Zekur.nl, targeting the young and healthy, contracts with only one hospital in most provinces, typically located in university cities.

[11] Similarly, in the United States, care contracted through Health Maintenance Organizations is more restricted in provider choice than in the form of Preferred Provider Organizations. Bardey and Rochet (2010) model competition among such health organizations where screening occurs through the diversity in contracted providers.

[12] See, for example, van Barneveld et al. (2001). For a discussion on the benefits and disadvantages on *ex post* risk adjustment in the US context, see, for example, Swartz (2003); Dow, Fulton, and Baicker (2010).

[13] Such variables depending on previous years' expenditure may, of course, reintroduce some of the perverse incentives associated with *ex post* risk adjustment.

[14] Variables to be included in *ex ante* risk adjustment must be readily available for the risk adjustment system to work well. In addition, if insurers can affect some variable, it is sometimes argued that these variables should not be included in the risk adjustment equation (Schokkaert, Dhaene, and Van de Voorde, 1998; Van de Ven and Ellis, 2000).

competing health insurers offer contracts to consumers. Consumers then choose which contract to accept. Consumers differ in expected health costs (high or low) and have heterogeneous preferences over the insurers. In this model, we study how market outcomes are affected by the risk adjustment scheme, and we investigate how the planner should choose its risk adjustment parameters if it wants to optimize a (weighted) welfare function.

We assume that the planner and the insurers can perfectly observe each customer's risk type. However, insurers are prevented by law to act upon this information, that is, we assume mandatory community rating. In other words, explicit price discrimination by insurers is not allowed. The insurers can use second-degree price discrimination to separate consumers, screening the high from the low types by use of the coverage parameter q offered by the contract. That is, an insurer cannot offer the same contract at different prices to different consumers. However, an insurer can offer two separate contracts (q^l, p^l), (q^h, p^h), each of which will be attractive to a different consumer type.

Along the lines of the *ex ante* risk adjustment systems observed in practice and described above, the planner can set risk adjustment subsidies $\rho^{l,h}$ (or taxes, when negative) to the insurer that depend only on the risk types of its insured.

The timing then is as follows.

(i) The planner sets risk adjustment parameters ρ^i, where $i = l, h$.
(ii) Insurers I_a and I_b each offer a menu of insurance contracts consisting of coverage q and price p, (q^i_a, p^i_a) and (q^i_b, p^i_b), respectively, where $i = l, h$.
(iii) Consumers of type $i = l, h$ choose which contract to accept from which insurer.
(iv) Risk adjustment payments are made: insurers receive ρ^i for each contracted consumer of type i from the planner.

After this, some insured fall ill and claim (part of their) medical expenses from the insurers.

We next discuss in detail the problems to be solved by the three groups of players: consumers, insurers, and planner.

□ **Contracts and consumers.** In modelling the interaction between insurers and consumers, we follow the setup in Rothschild and Stiglitz (1976). We consider two types of consumers, h and l, where the fraction of h types is denoted by $\lambda \in \langle 0, 1\rangle$. A consumer of type $i = h, l$ has expected medical costs θ^i, with $\theta^h > \theta^l > 0$. An insurance contract consists of a price p^i and a coverage q^i, implying that an agent has to pay a fraction $1 - q^i$ of medical costs herself. Alternatively, q can be interpreted as the generosity of the insurance contract. This could take the form of copayments and deductibles but also to what degree the insured is free to choose the provider she wishes.[15]

For analytical tractability, we assume a mean-variance utility function: a consumer of type i who buys an insurance contract with price p^i and coverage q^i has utility

$$u^i = w - (1 - q^i)\theta^i - p^i - \frac{1}{2}r\sigma^2(1 - q^i)^2. \tag{1}$$

Here, w denotes the initial wealth of the agent, the variance of a consumer's expected medical costs equals σ^2, and $r > 0$ measures the agent's degree of risk aversion.[16] An added advantage of mean-variance utility is that we can neatly separate efficiency and solidarity. As we shall see, we model solidarity by increasing the weight of the high type in the planner's objective function. When risk aversion is modelled using a utility function concave in p, the high type (facing a higher premium p) has a higher marginal utility of income. This immediately gives an incentive

[15] See, for example, Chernew and Frick (1999) and Bardey and Rochet (2010), who model contract generosity in terms of the degree of restrictions in provider choice in a managed care context.

[16] Note that utility is assumed linear in price p^i. This can be consistent with global risk aversion, if health shocks are large but have very low probability. Then, expected costs θ^i and p^i will be low, making a linear approximation adequate, but variance over health shocks is much higher (see Bardey and Rochet, 2010).

for the planner to redistribute from the low-risk to the high-risk type, making it harder to separate efficiency and solidarity considerations.

We consider only insurance contracts with $q \in [0, 1]$. Contracts with $q > 1$ are ruled out as these would invite serious moral hazard problems: a consumer could then earn money by undergoing treatment. Similarly, $q < 0$ is ruled out because consumers would not report any treatment on which they spent money.

As a result of the community rating requirement, consumers can choose among all contracts on offer. To ensure that in equilibrium a consumer chooses the contract intended for its type, we focus on contracts that respect the incentive compatibility constraints for the h-type and the l-type

$$IC_h : u^h \geq u^l - \Delta\theta(1 - q^l) \tag{2a}$$

$$IC_l : u^l \geq u^h + \Delta\theta(1 - q^h), \tag{2b}$$

where $\Delta\theta = \theta^h - \theta^l > 0$. It follows from adding the two incentive compatibility constraints that $q^h \geq q^l$. The constraint for the low type then implies that $p^h \geq p^l$.

Apart from the distinction in risk types, consumers also differ horizontally in preferences for insurers. We model this in Hotelling fashion. The two insurers I_a and I_b are located on either end of a Hotelling beach of length 1, with a mass one of consumers distributed along the beach. Consumers located at point $x \in [0, 1]$ incur travel cost tx (or $t(1 - x)$) when choosing for a contract from I_a (or I_b). To bring out the correlation between risk type and switching elasticity that we observe in practice, we assume that the two types of consumers have different distributions over the beach. The density function on [0, 1], which for each type is symmetric around $x = \frac{1}{2}$, is given by $f(x)$ for an h-type and by $g(x)$ for an l-type, with respective cumulative distribution functions $F(x)$ and $G(x)$. We model the idea that the h-types are less elastic in switching insurer by assuming that F has relatively more mass at the extremes and less in the middle. For G, it is the other way around. Consequently, l-types are more willing to buy from any insurer, whereas the h-types are “biased” toward the insurer close to them. Thus, we assume that the cumulative distribution functions $F(x)$ and $G(x)$ satisfy

$$\frac{f(\frac{1}{2})}{F(\frac{1}{2})} < \frac{g(\frac{1}{2})}{G(\frac{1}{2})}, \tag{3}$$

where $F(\frac{1}{2}) = G(\frac{1}{2}) = \frac{1}{2}$.

When insurer I_a offers a contract giving utility u_a^h to the high types, and I_b offers utility u_b^h, the indifferent high-type consumer is located at $\frac{1}{2} + \frac{u_a^h - u_b^h}{2t}$ and firm I_a's market share in the θ^h market is given by

$$F\left(\frac{1}{2} + \frac{u_a^h - u_b^h}{2t}\right). \tag{4}$$

Consequently, if both insurers offer the same utility the market is split fifty-fifty. Similar arguments hold for the low types.

□ **Insurer competition.** Insurers I_a and I_b compete in menus of contracts to maximize their profits, in the spirit of Stole (1995). We assume that the whole market is served and that the relevant outside option for an agent is switching to the competing insurance company. This is called “full scale competition” (Schmidt-Mohr and Villas-Boas, 1999) or “pure competition” (Stole, 1995). This assumption is justified if health insurance is mandatory or if consumer travel costs are low enough. If I_a offers higher utility for a given utility offered by I_b, it gains market share from I_b. The lower t is, the faster this goes. Hence, lower t corresponds to more intense competition and perfect competition corresponds to $t = 0$.

In determining their contract offers, insurers will not only take into account the direct profits from consumers, but also the contribution ρ^i that they will get from the risk adjustment scheme for every θ^i-type consumer they contract with.

We find it convenient to write an insurer's per-consumer profit margin π^i for type $i = l, h$, that is, price minus expected cost plus the risk adjustment contribution, in terms of the utility offered to the consumer, u^i, and the coverage parameter q^i, using equation (1):

$$\pi^i(u^i, q^i) = p^i - q^i\theta^i + \rho^i = w - u^i - \theta^i + \rho^i - \tfrac{1}{2}r\sigma^2(1 - q^i)^2. \tag{5}$$

We see that the risk adjustment ρ^i affects market outcomes by changing the margin that health care insurers make on different types of consumers.

We will look for a symmetric equilibrium where the high-type incentive constraint binds. In that case, we can write insurer I_a's problem as follows:

$$\max_{u_a^h, u_a^l, q_a^h, q_a^l} \lambda F\left(\frac{1}{2} + \frac{u_a^h - u_b^h}{2t}\right)\pi^h(u_a^h, q_a^h) \;+ (1-\lambda)G\left(\frac{1}{2} + \frac{u_a^l - u_b^l}{2t}\right)\pi^l(u_a^l, q_a^l)$$
$$+\mu(u_a^h - u_a^l + \Delta\theta(1 - q_a^l)), \tag{6}$$

subject to $q_a^h, q_a^l \in [0, 1]$ and with a similar expression for insurer I_b. μ denotes the shadow price of the incentive compatibility constraint. Our candidate symmetric Nash equilibrium then satisfies the following first-order conditions:

$$q^h = 1 \tag{7a}$$

$$-\frac{\lambda}{2} + \frac{\lambda}{2t}f(\frac{1}{2})(w - \theta^h + \rho^h - u^h) + \mu = 0 \tag{7b}$$

$$-\frac{1-\lambda}{2} + \frac{1-\lambda}{2t}g(\frac{1}{2})(w - \theta^l + \rho^l - \frac{1}{2}r\sigma^2(1 - q^l)^2 - u^l) - \mu = 0 \tag{7c}$$

$$\frac{1-\lambda}{2}r\sigma^2(1 - q^l) = \mu\Delta\theta. \tag{7d}$$

We note that the solutions to equations (7a) – (7d) do not necessarily constitute a solution to the full problem for two reasons. First, objective (6) might not be concave. Second, given that the other insurer offers the candidate a symmetric equilibrium contract, profitable deviations might exist in which the high-type incentive constraint does not bind. In Appendix B, we explore sufficient conditions for the symmetric solution to be a Nash equilibrium.[17] In the remainder of this article, we will assume such conditions to hold.

From (7a), we immediately get that the h-type receives full insurance. Coverage for the low type is given by

$$1 - q^l = \frac{\mu\Delta\theta}{\frac{1}{2}(1-\lambda)r\sigma^2}. \tag{8}$$

Hence, the lower the shadow price μ, that is, the less tightly the incentive compatibility constraint binds, the higher the coverage for the low type.

□ **The planner's objective.** The planner chooses risk adjustment ρ^h, ρ^l to improve the outcome in the insurance market. We assume that the scheme is self-financing. The balanced budget equation for the planner can be written as

$$\lambda\rho^h + (1-\lambda)\rho^l = 0. \tag{9}$$

[17] See Olivella and Vera-Hernández (2007) for a numerical analysis of the case with corner solutions.

We introduce a general objective function for the planner. The planner chooses $\rho^{l,h}$ to maximize[18]

$$W = (1 - \omega)u^l + \omega u^h + \gamma(\pi_a + \pi_b), \tag{10}$$

subject to budget balance, and taking into account the insurance market equilibrium that will result from the choice of ρ^i, characterized by the first-order equations (7a)–(7d).

If $\omega = \lambda$ and $\gamma = 1$, the objective function corresponds to total welfare. Total welfare as objective function is standard in the analysis of risk adjustment (see, for instance, Glazer and McGuire, 2002; Jack, 2006). However, it is not obvious that this is the appropriate analysis in a context with imperfect competition. Competition authorities and regulators around the world explicitly claim that their objective is consumer welfare, not total welfare. Further, in the health care sector, policy motivations for interventions such as risk adjustment usually focus on consumers and solidarity between different types of consumers, rather than on efficiency.[19]

We capture the planner's focus on consumers by changing the weights γ and ω in the objective function. Lower γ means that the planner puts higher value on consumer surplus than on profits, as in Baron and Myerson (1982). With $\gamma = 0$, the planner maximizes consumer welfare (and does not value insurers' profits). Solidarity with high-risk agents can be captured by $\omega > \lambda$: the planner gives higher weight to (unlucky) high-risk people than their fraction in the population.[20] This makes (in)equity a concern besides efficiency. Fleurbaey and Schokkaert (2011) give an overview of why policy makers care about equity in health care.

In this article, we are interested in two questions. First, what outcome does the planner implement? For instance, does the planner want both insurance contracts to have efficient coverage, $q^{h,l} = 1$? The second question is: for a given optimal coverage $q^{h,l}$, what does the risk adjustment $\rho^{h,l}$ look like? For instance, does the difference in risk adjustment contributions equal the difference in expected costs?

4 Optimal coverage

■ In this section, we concentrate on finding the optimal coverage for low-cost types. We show here that when the planner puts a higher weight on consumers than on insurers ($\gamma < 1$), it may be optimal to choose to implement a positive distortion. In contrast, conventional risk adjustment aims at removing all incentives for selection, setting the equilibrium distortion $1 - q^l$ equal to zero. In the next section, we show how the planner can implement the optimal coverage: what set of risk adjustment contributions $\rho^{h,l}$ brings about an insurance market equilibrium with the desired coverage.

We analyze optimal coverage in terms of the optimal distortion $1 - q^l$.[21] To do so, we eliminate $\rho^{h,l}$ by combining the first-order conditions for $u^{h,l}$, equations (7b) and (7c), with the budget balance condition (9).[22] We can then solve for u^h, u^l in terms of $1 - q^l$ by using the binding incentive compatibility condition for the high types (2a), and by substituting for μ using equation (8). The resulting expressions for u^h, u^l, combined with the relation between profits and utilities, equation (5), allow us to express industry profits $\Pi = \lambda\pi^h + (1 - \lambda)\pi^l$ as a function of $1 - q^l$.

[18] Note that we focus on symmetric equilibria. Hence, the (dis)utility a consumer experiences when buying one brand rather than the other, the travel costs in our Hotelling setup, does not change in the comparative static exercises that we do below. Therefore, we ignore this term in the planner's objective function.

[19] Relatedly, in his overview article, Ellis (2008) mentions "the concept of 'optimal risk adjustment' in which the sponsor's goal is to maximize consumer welfare."

[20] This difference in weighting could reflect purely normative goals on the part of the planner. Alternatively, from a positive perspective, the different welfare weights correspond to the set of "interim efficient" equilibria as explained in Holmström and Myerson (1983) and Ledyard and Palfrey (1999).

[21] Note that, by the first-order conditions, for any choice of risk adjustment $\rho^{h,l}$, high-cost types are always fully insured in equilibrium, $q^h = 1$.

[22] See the proof of Proposition 1 for the explicit derivation.

Putting together these results, we find the following marginal effects of the distortion $1 - q^l$ on u^h, u^l and Π:

$$\frac{du^h}{d(1-q^l)} = -(1-\lambda)r\sigma^2(1-q^l) + (1-\lambda)r\sigma^2 ME - (1-\lambda)\Delta\theta \tag{11a}$$

$$\frac{du^l}{d(1-q^l)} = -(1-\lambda)r\sigma^2(1-q^l) + (1-\lambda)r\sigma^2 ME + \lambda\Delta\theta \tag{11b}$$

$$\frac{d\Pi}{d(1-q^l)} = -(1-\lambda)r\sigma^2 ME, \tag{11c}$$

where we defined the margin effect ME as

$$ME = \frac{\frac{t}{f(\frac{1}{2})} - \frac{t}{g(\frac{1}{2})}}{\Delta\theta} > 0,$$

with the inequality following from the assumption on the difference in elasticities, equation (3).

For the interpretation of these expressions, consider first equations (11a) and (11b). We can distinguish three effects of increasing the distortion $1 - q^l$. The first two are common to both types' utilities, whereas the third one measures the difference between u^h and u^l.

The first term represents the *distortion effect*. As we can see from the expression for the low type's utility, equation (1), the utility loss as a direct result of the distortion itself is proportional to $\frac{1}{2}(1 - q^l)^2$ and hence, the derivative is proportional to $1 - q^l$. With $1 - q^l = 0$, the distortion effect is minimized. The distortion affects u^l directly, but via the binding incentive compatibility condition (2a), u^h is affected as well.

The second term is the *margin effect*: if we increase the distortion $1 - q^l$, insurers compete more fiercely, driving down prices and increasing consumer utility. The reason is that, by the difference in elasticities, the low type is more likely to switch insurer than the high type. Without risk adjustment, insurers therefore compete more fiercely on the low type than the high-type segment. Introducing a risk adjustment that reduces the expected cost difference between the two types involves subsidizing the high types and penalizing the low types. As a result, the margin on the low types is reduced, and so are the insurers' incentives to compete for these types. Although such risk adjustments reduce the selection incentives for insurers, and the equilibrium distortion that comes with selection, they therefore also mute competition. Hence, a reduction in the distortion goes hand-in-hand with a reduction in the intensity of competition.

The final term is the *inequality effect*. It measures the difference between low-type and high-type utilities. As can be seen from the binding incentive compatibility condition for the high types, equation (2a), the difference between u^l and u^h increases with $1 - q^l$. Given the distortion and the margin effect, an increase in $1 - q^l$ raises u^l but reduces u^h such that $d(u^l - u^h)/d(1 - q^l) = \Delta\theta$.

Insurer profits are only sensitive to the margin effect. As explained above, an increase in distortion $1 - q^l$ goes hand-in-hand with a stronger incentive to compete in the elastic low-type market. Hence, competition for these types increases with $1 - q^l$, leading to a decrease in margins on that segment. However, as contracts have to respect incentive compatibility, high–type prices are also dragged down with the prices for low types. It is this effect that we call "competition leverage": by intensifying competition on the low-type segment (and raising $1 - q^l$), insurers offer higher values of u^l; via the incentive compatibility constraint, u^h goes up as well. Competition is leveraged from the low- to the high-type segment.

Taking these separate effects on consumer utilities and profits together, we can derive the effect of $1 - q^l$ on the planner's objective W, and characterize how the optimal coverage q^l depends on the weights γ and ω.

Proposition 1. The effect of $1 - q^l$ on W is given by

$$\frac{dW}{d(1-q^l)} = (1-\lambda)r\sigma^2\left(-(1-q^l) - \frac{\omega-\lambda}{1-\lambda}\frac{\Delta\theta}{r\sigma^2} + (1-\gamma)ME\right). \tag{12}$$

Hence,

(i) with $\gamma = 1$, it is optimal to have full coverage for both types, $q^l = q^h = 1$,
(ii) with $\gamma < 1$ and $\omega = \lambda$, the planner optimally sets $q^l < 1$,
(iii) if $(1-\gamma)ME > \frac{\Delta\theta}{r\sigma^2}$, then $q^l < 1$ for each $\omega \in [\lambda, 1]$.

Looking at the marginal effect of the distortion on W, the first effect is the distortion effect. As the welfare loss is second-order at first-best coverage $q^l = 1$, the distortion effect vanishes for this value. The second effect is the inequality effect. Because the inequality effect raises u^l by as much as it reduces u^h, it disappears if the planner focuses on total consumer welfare ($\omega = \lambda$). If the planner gives higher weight to high types, that is, $\omega > \lambda$, the inequality effect negatively affects W. The third effect is the margin effect. It raises utility for both types but reduces profits. The margin effect disappears if the planner maximizes total welfare ($\gamma = 1$) but increases W if the planner gives lower weight to insurers ($\gamma < 1$).

If $\gamma = 1$ and $\omega = \lambda$, so that the planner maximizes (unweighted) total welfare, equation (12) shows that the optimal $q^l = 1$, that is, no distortion. By our assumption of full scale competition, every consumer buys insurance. Hence, the insurers' profit margins are just a transfer from consumers to insurers, which drops out of total welfare. Further, $\omega \geq \lambda$ implies that increases in $u^l - u^h$ do not raise W.[23] Hence, we get $1 - q^l = 0$.

If we give lower weight to the insurers' profits than to consumer utility ($\gamma < 1$), the margin effect starts to play a role: leveraging competition becomes attractive. The planner distorts coverage for the low types to reduce profits and increase utilities. Thus, $\gamma < 1$ and $\omega = \lambda$ implies that $1 - q^l > 0$ is optimal.

Finally, increasing $\omega > \lambda$, makes it less attractive to raise $1 - q^l$. As we put more weight on the high types, the inequality effect (measuring the increase in $u^l - u^h$) kicks in, and from that a larger distortion will decrease W. However, if even for $\omega = 1$ we have that $dW/d(1-q^l) > 0$ (evaluated at $q^l = 1$), this expression is positive for all $\omega \in [\lambda, 1]$. This is the case if $(1-\gamma)ME > \frac{\Delta\theta}{r\sigma^2}$. If increasing $1 - q^l$ raises W in the case $\omega = 1$, it follows from equations (11a) and (11b) above that both u^h and u^l increase with the distortion. In this case, the margin effect is so strong, that the price reduction from leveraging competition from the low to the high segment dominates the distortion and inequality effects even for the high types. Note that the margin effect is stronger the less competitive the insurance market is (high t), the bigger the difference in elasticities is (driven by $g(\frac{1}{2}) - f(\frac{1}{2})$), and the lower the difference in expected costs $\Delta\theta$.

We remark, finally, that in the absence of a margin effect ($ME = 0$)–as the existing risk adjustment literature assumes–the proposition implies that efficient insurance ($q^l = 1$) maximizes W for each $\omega \in [\lambda, 1]$. In this sense, the existing literature is correct in focusing on first best. However, as discussed in the Introduction, it is well documented that h-types are less likely to switch insurer compared to l-types. This difference in switching behavior implies that $ME > 0$ and therefore that efficiency and weighted welfare W can diverge.

5 Optimal risk adjustment

■ Above, we derived the optimal distortion that a planner would like to introduce. Here, we consider the risk adjustment contributions, $\rho^{h,l}$, that implement a given distortion $1 - q^l$. We will show that, in general, these optimal risk adjustment contributions require more than merely compensating for differences in expected costs.

[23] Although not customary in the health economics literature, one can consider $\omega < \lambda$ to trace out the interim efficient outcomes: see footnote 20. Then, to bias the outcome in favor of the low type, the planner implements $q^l < 1$.

Perhaps surprisingly, this is true even in the case where the planner wants to eliminate all selection incentives and to obtain $q^l = q^h = 1$. As shown in Proposition 1, this is optimal if the planner maximizes total welfare. As discussed above, the existing risk adjustment literature focuses on equalizing expected costs to achieve this goal. At first sight, this seems equivalent to equalizing price cost margins. As the h-type is less elastic in switching insurer than the l-type, the markup is higher for the h-type. Naively, one might expect that the higher markup reduces the compensation to be paid for higher-cost consumers to make insurers indifferent. Efficient risk adjustment would then undershoot the cost difference: $\rho^h - \rho^l < \theta^h - \theta^l$. However, the next result shows that this intuition is incorrect.

Proposition 2. The risk adjustment contributions ρ^h, ρ^l that implement first best ($q^{h,l} = 1$) satisfy

$$\rho^h - \rho^l = \left(\theta^h - \theta^l\right) + \left(\frac{t}{f(\frac{1}{2})} - \frac{t}{g(\frac{1}{2})}\right).$$

That implies $\rho^h > \rho^l$, that is, risk adjustment is in the standard direction, and $\rho^h - \rho^l > \theta^h - \theta^l$, that is, risk adjustment has to overshoot the difference in expected costs, correcting for the difference in markups.

The intuition for this result is the following. With efficient coverage, $q^{l,h} = 1$, contracts for both types are the same. Thus, prices for both contracts have to be the same, p. As the incentive compatibility constraint is slack with full coverage (we have $\mu = 0$), each contract should also separately optimize the insurer's profits, and hence, markups for contracts $i = l, h$ satisfy the standard Lerner equation

$$p - C^i = \frac{p}{\eta^i}, \tag{13}$$

where η^i is type i's effective demand elasticity[24] and costs C^i are the insurer's effective costs, including the risk adjustment contribution, $C^i = \theta^i - \rho^i$. In our duopoly model, residual demand elasticity is governed by consumers' tendencies to switch insurer, which is higher for low types than for high types. For equation (13) to hold for both types at the same price, effective costs C^i should not be equalized, but should also reflect this difference in markups: in offering equal contracts to both types, insurers also face the opportunity cost of the higher markup on the h-type, $\frac{t}{f} - \frac{t}{g} > 0$. If this opportunity cost is overlooked (such that $\rho^h - \rho^l = \theta^h - \theta^l$), insurers have an incentive to separate the types by extracting rents from the inelastic h-type and competing more vigorously for the elastic l-type, leading to $q^l < 1$.

The extent to which risk adjustment has to overshoot the cost difference is increasing in insurers' market power (t) and in the difference in demand elasticities ($g(\frac{1}{2}) - f(\frac{1}{2})$). If either the market is perfectly competitive ($t = 0$) or demand elasticities are the same for both types, there is no need to overshoot the cost difference.

In other words, the current risk adjustment models with their exclusive focus on the cost/supply side will not restore efficiency in health insurance markets where insurers have market power and types differ in their tendency to switch insurer. Even if one would perfectly compensate for cost differences between types, ignoring their differing demand elasticities leads to selection behavior by insurers and, hence, inefficient health insurance.

If risk adjustment corrects for both the difference in costs and the difference in markups, and consequently overshoots the standard risk adjustment ($\rho^h - \rho^l > \theta^h - \theta^l$), we find $q^l = 1$. We call this *first-best* or *efficient risk adjustment*. Efficient risk adjustment follows if the insurer can

[24] Satisfying $\frac{p}{\eta^h} = \frac{t}{f(\frac{1}{2})}$, $\frac{p}{\eta^l} = \frac{t}{g(\frac{1}{2})}$.

optimize the contracts of each type separately without worrying about incentive compatibility, and the resulting contracts will satisfy IC.[25]

Proposition 2 gives the risk adjustment contributions for the case where the planner wants to implement the efficient outcome $q^l = 1$. However, Proposition 1 also derives conditions under which the planner's objective W is maximized at $q^l < 1$. To explore the implications for the risk adjustment contributions, we first note that $1 - q^l$ is increasing in ρ^l:

Lemma 1. The distortion in low-type coverage is increasing with the low-type risk adjustment contribution,

$$\frac{d(1 - q^l)}{d\rho^l} > 0. \tag{14}$$

Equivalently, this states that standard risk adjustment decreases the distortion in low-type coverage: a lower contribution to the low types ρ^l, and hence, a higher contribution for the high types ρ^h (as a result of budget balance), lowers the incentives to select, and hence, lowers $1 - q^l$.

Combining Lemma 1 with Proposition 1, we see that ρ^l should be higher (or less negative), the lower the planner's weight γ on insurers' profits and the higher the margin effect ME. As increasing ρ^l reduces profits, the planner is only willing to do this if the weight on profits is relatively small. Moreover, competition leverage's effect on welfare is higher as the difference in elasticities is bigger.

The surprising observation now is that from an increase in ρ^l, and hence, to keep budget balance, an associated decrease in ρ^h, high types may benefit. Although the reduction in subsidy ρ^h raises the costs for the insurer of insuring the high types, the competition leverage effect can be so strong that maximizing u^h implies $\rho^l > 0 > \rho^h$, or in other words, risk adjustment in the "wrong" direction.

Proposition 3. For $\Delta\theta$ small enough, u^h is maximized by setting $\rho^h < 0$. *A fortiori,* consumer welfare is then optimized for $\rho^h < 0$ for any weight ω.

The condition that $\Delta\theta$ should be small follows immediately from Lemma 1 and Proposition 1 with $\omega = 1$. When maximizing u^h, $1 - q^l$ will be high if the costs of increasing $1 - q^l$ are small compared to the benefits. That is, the distortion and inequality effects need to be small compared to the margin effect, ME. If $\Delta\theta$ is small, the inequality effect becomes small, whereas at the same time, for $\Delta\theta$ close to zero, ME becomes big. If $\Delta\theta$ is small enough, ME also outweighs the distortion effect, and ρ^h actually turns negative.

Finally, of course, if high–type consumers win from having negative risk adjustment ρ^h, then low types will certainly be better off. As a result, if the difference in demand elasticities is sufficiently high (so ME is large), and cost differences are not too big, all consumers are better off without risk adjustment than with first-best risk adjustment. Therefore, even though from an efficiency point of view risk adjustment is desirable, a consumer-oriented planner may want to abstain from it.

6 Mandatory pooling

■ As shown in Proposition 2, first-best risk adjustment implies that insurers offer both types the same contract. This might suggest that there is no role for mandatory pooling, as long as risk adjustment is chosen optimally. In this section, we show that this intuition is not correct in general. When the planner puts more weight on consumers, $\gamma < 1$, mandating pooling generally leads to

[25] In the equilibrium in Proposition 2, insurance companies offer the same contract to low and high types, so we effectively have a pooling equilibrium. Making pooling mandatory will not affect the outcome in this case. Yet, as shown in Section 6, mandatory pooling is not an irrelevant instrument for the planner.

higher W than allowing second-degree price discrimination (separating contracts). Moreover, in many cases the W-maximizing insurance contract features inefficient insurance ($q < 1$) for both types.

With mandatory pooling, each insurer is allowed to offer only a single contract with some price p and a copayment $1 - q$. Hence, the insurer cannot price discriminate between types. The margin π^i on type $i = l, h$ therefore equals

$$\pi^i = p - q\theta^i + \rho^i.$$

Given a price and a copayment, the utility of the h-type satisfies

$$p = w - u^h - (1-q)\theta^h - \frac{1}{2}r\sigma^2(1-q)^2, \tag{15}$$

and equation (2a) holding with equality determines the utility that the l-type receives given the utility of the h-type. However, the latter is not an IC constraint here, as there is only one contract, but an "accounting identity," which relates the utilities of the h-type and the l-type consumer. Put differently, this equation ensures that $p^h = p^l = p$ when $q^h = q^l = q$.

Hence, insurer I_a's optimization problem can be written as

$$\max_{u_a^h, u_a^l, q_a} \lambda F\left(\frac{1}{2} + \frac{u_a^h - u_b^h}{2t}\right)\pi^h(u_a^h, q_a) \; + (1-\lambda)G\left(\frac{1}{2} + \frac{u_a^l - u_b^l}{2t}\right)\pi^l(u_a^l, q_a)$$
$$+\mu(u_a^h - u_a^l + \Delta\theta(1-q_a)). \tag{16}$$

The optimization problem is similar to equation (6), with the additional requirement that $q^l = q^h$. The first-order conditions for q, u^h, and u^l in symmetric equilibrium can be written as

$$\frac{1}{2}r\sigma^2(1-q) = \mu\Delta\theta \tag{17a}$$

$$-\frac{\lambda}{2} + \frac{\lambda}{2t}f\left(\frac{1}{2}\right)(w - \theta^h + \rho^h - \frac{1}{2}r\sigma^2(1-q)^2 - u^h) + \mu = 0 \tag{17b}$$

$$-\frac{1-\lambda}{2} + \frac{1-\lambda}{2t}g\left(\frac{1}{2}\right)(w - \theta^l + \rho^l - \frac{1}{2}r\sigma^2(1-q)^2 - u^l) - \mu = 0. \tag{17c}$$

We can again analyze the welfare-maximizing value of the distortion $1 - q$, a distortion which is in this case common to both types. We proceed along the lines of Section 4 to find the effect of coverage on the planner's objective function W. Eliminating the risk adjustments, using budget balance and the first-order conditions 17b, 17c, and using the expression for the shadow price μ from equation (17a) as well as the accounting identity equation (2a) relating u^h to u^l, we arrive at the following result.

Proposition 4. Under mandatory pooling, the effect of the distortion $1 - q$ on W is given by

$$\frac{dW}{d(1-q)} = r\sigma^2\left(-(1-q) - (\omega - \lambda)\frac{\Delta\theta}{r\sigma^2} + (1-\gamma)ME\right). \tag{18}$$

Hence,

(i) with $\gamma = 1$, it is optimal to have full coverage for both types, $q = 1$,
(ii) with $\gamma < 1$ and $\omega = \lambda$, the planner optimally sets $q < 1$,
(iii) if $(1-\gamma)ME > (1-\lambda)\frac{\Delta\theta}{r\sigma^2}$, then the planner sets $q < 1$ for all $\omega \geq \lambda$.

Comparing with the case of separating contracts analyzed in Proposition 1, we see the following.The first term, the distortion effect, is bigger. With mandatory pooling, both types get distorted coverage, rather than only the $1 - \lambda$ low types. Hence, the costs of a marginal increase of $1 - q$ is larger. However, starting from first best, this effect equals 0. The inequality effect, in contrast, remains equal. The gap between high, and low–type utilities, captured by this second term, remains governed by the same condition (2a).

The final term, the (positive) margin effect, is larger under mandatory pooling. The intuition for this is as follows. As the planner increases ρ^l, insurers compete more fiercely for l-types, thereby reducing p^l and increasing u^l. The IC constraint (2a) then implies that either u^h increases as the planner intended, or coverage q falls because this is the insurers' instrument to keep u^h low. When second-degree price discrimination is allowed, reducing coverage q^l is relatively cheap for insurers. Indeed, q^l affects only profits on the l-market. However, with mandatory pooling, reducing q affects profits on both the l- and h-market segment. This makes it expensive for insurers to use q as a "defense mechanism" against competition leverage. Put differently, for a given distortion $1 - q$, mandatory pooling allows for a bigger value of ρ^l and hence, more intense competition on the l-segment of the market. Consequently, the distortion has a bigger positive effect on welfare under mandatory pooling as it induces more competition leverage.

We can combine these observations to compare the planner's optimum under mandatory pooling and separating contracts.

Proposition 5. In the optimum under mandatory pooling:

(i) the distortion $1 - q^{pool}$ is greater than with separating contracts, $q^{pool} \leq q^{sep}$, with strict inequality unless $\gamma = 1$ or $\omega = \lambda$;
(ii) the planner's objective is larger than with separating contracts, $W^{pool} \geq W^{sep}$, with strict inequality if $q^{pool} < 1$.

First, we have seen that under mandatory pooling, the inequality effect is relatively less important, compared to the distortion and margin effects, than in the previous case. When $\omega = \lambda$, the inequality effect vanishes and the resulting distortion is no different from the separating case. When $\omega > \lambda$, with mandatory pooling, it will be easier to satisfy the condition that benefits of the margin effect outweigh the costs of the inequality effect than in the separating case. Hence, we obtain $q < 1$ in the optimum for a larger range of parameters than before. More generally, the larger contribution from the margin effect implies that the optimal distortion will then be greater.

Second, because in the first-best, no-distortion case, welfare under pooling and separation coincide, and as the marginal effect of increasing the distortion $1 - q$ is greater with mandatory pooling, whenever we have a positive distortion, welfare will be higher in the pooling case. In particular, we have that consumers are better off under mandatory pooling than under separation.

If, in addition to enforcing the same q on each contract, the planner can enforce the level of q, welfare can be further enhanced. By fixing q (at its efficient level), insurers can no longer use reductions in q as a defense against competition leverage. Hence, increasing ρ^l leads to more intense competition on the l-market, which then feeds one-to-one to the h-market as well.

In practice, it is unlikely that the government can completely fix q. If q is interpreted as the coinsurance rate, then indeed, it can be fixed. However, if q represents other dimensions of generosity, such as responsiveness to queries by the insured, this is much less likely to be contractible by the government.

However, the more general point is that reducing insurers' instruments to select types enhances the effect of competition leverage. One can think here of the Dutch government determining the coverage of the basic insurance package, instead of leaving this to insurers.

7 Imperfect observation of risk

■ Above, we assume that the planner can *ex post* perfectly observe each customer's type. In reality, the planner may only have a noisy signal of a customer's θ. The question is: are the results above robust to the imperfect observation of risk?

Following Glazer and McGuire (2000), let us assume that the planner cannot perfectly observe the consumers' types. Rather, the planner observes an imperfect signal (H, L) of consumers' types (h, l). With probability P^h, the h-type will give a high signal H, and with probability $1 - P^h$, the high–type's signal will be L. Likewise, with probability P^l, the l-type's signal will be H (and with $1 - P^l$, it will be L). Of course, for the signals to be informative at all we require

$$0 \leq P^l \leq \frac{1}{2} \leq P^h \leq 1.$$

Thus the h-type will more likely produce the H-signal, and the l-type will more likely produce signal L.

The planner can now base the risk adjustment on only the observed signal (H, L). Consumers, however, know their types and base their choices on their actual types (h, l). Call the risk adjustment based on the signal $\rho^{H,L}$. Then, an insurer will get an effective subsidy on the high types it attracts equal to

$$\tilde{\rho}^h = P^h \rho^H + (1 - P^h)\rho^L,$$

and likewise the effective subsidy for the low types will equal

$$\tilde{\rho}^l = P^l \rho^H + (1 - P^l)\rho^L.$$

Note that the $\tilde{\rho}$'s interpolate between the ρ's.

To analyze the optimal $\rho^{H,L}$ in this case, note that our analysis when signals are perfect fully carries through with ρ replaced by $\tilde{\rho}$. Consequently, the optimal subsidies that we found in that case are the optimal effective subsidies $\tilde{\rho}$. The subsidies the planner should give based on the signals (H, L) should therefore be an exaggeration of the optimal subsidies, with positive subsidies being increased and negative ones being decreased. This observation is independent of whether the planner enforces mandatory pooling or not. Hence, we get in both cases:

$$\rho^H = \frac{1}{P^h - P^l}\left((1 - P^l)\tilde{\rho}^h - (1 - P^h)\tilde{\rho}^l\right), \tag{19a}$$

$$\rho^L = \frac{1}{P^h - P^l}\left(-P^l\tilde{\rho}^h + P^h\tilde{\rho}^l\right). \tag{19b}$$

To illustrate, when it is optimal to subsidize the low types ($\tilde{\rho}^l > 0$) at the expense of the high types (Proposition 3), imperfect observation of types only makes this stronger in the sense that $\rho^H < \tilde{\rho}^h < 0 < \tilde{\rho}^l < \rho^L$. We conclude that our results are, indeed, robust to the introduction of noisy signals about consumers' types.

Traditional risk adjustment uses variables correlated with expected costs as (imperfect) signals of agents' types. However, given that elasticities differ, the observation that someone switches is also an imperfect signal of type. Suppose the planner wants to target risk adjustment subsidies at elastic l-types. Then, the planner can pay an insurer relatively more for new customers (which, by switching [imperfectly] signal to be l-type) than for existing customers. Of course, community rating would still be enforced: insurers are not allowed to offer different contracts to new customers and existing customers.

8 Conclusion

■ In several countries, imperfect competition in markets for health care insurance goes hand-in-hand with risk adjustment. The resources allocated to risk adjustment are significant; for example, approximately 3% of GDP in the Netherlands. Conventional risk adjustment focuses on the supply side and tries to correct as best as possible for cost differences between high-cost and low-cost consumers. Existing theoretical articles argue that this improves efficiency by reducing insurers' selection incentives, whereas at the same time guaranteeing access to health insurance for high-risk consumers by lowering insurance premiums.

We study the consequences of the demand side for optimal risk adjustment. In particular, we take into account that consumers' price elasticities are negatively correlated with their expected health care costs. We find that the difference in switching behavior creates a trade-off between efficiency and consumer welfare. From an efficiency point of view, to nullify selection incentives, risk adjustment needs to overcompensate the high-risk types (relative to their costs) to get an efficient outcome in the insurance market. If the planner places more weight on consumer welfare, reducing the level of risk adjustment, and as a consequence allowing some distortion in coverage, increases consumer welfare by leveraging competition from the elastic low-risk market to the less elastic high-risk market. Mandatory pooling can increase consumer surplus even further, at the cost of efficiency.

Our analysis has three main policy implications. First, an exclusive focus on increasing the accuracy of cost forecasts in refining risk adjustment systems is misguided. Risk adjustment should not focus on the costs, that is, the supply side, only. Efficiency cannot be fully restored in this way. Risk adjustment resources will be partially wasted by this focus on costs. Risk adjustment systems should take into account the impact they have on insurer competition, and the possibility of leveraging competition that arises because of the relation between consumer type and tendency to switch insurers. Our model shows how the literature studying this relation can be taken on board. We show how predicted costs and switching elasticities should be combined to find the risk adjustment transfers that maximize the planner's objective. Accounting for the demand side of health insurance may be easier than it seems, because the empirical literature studying consumers' brand sensitivity uses many of the same explanatory variables as the literature on risk adjustment. Our analysis also shows that nullifying selection incentives using risk adjustment does not benefit consumers. Thus, risk selection by insurers should not be viewed as bad per se, as seems the case in policy debates.

Second, the benefits of competition leverage increase as the types' utility levels are more closely tied together. With mandatory pooling, it is more costly for insurers to soften competition leverage by varying coverage. Other ways in which insurers can soften intra brand competition include differentiating insurance contracts by treatments covered, providers contracted, etc. The more the planner standardizes the health insurance contract, the more it can benefit from competition leverage.

Third, policy makers should be clear on their goals. When setting risk adjustment levels, policy makers should decide whether their goal is to optimize total welfare or consumer welfare. As we demonstrate, if everyone has the same tendency to switch insurers, efficiency, consumer welfare, and the utility of the high-risk types are all maximized by implementing the first best. Hence, the exact goal of the risk adjustment scheme is immaterial. However, in the empirically relevant situation, the switching elasticity is negatively correlated with expected health care costs. Although conventional risk adjustment most likely promotes efficiency, it will reduce consumer surplus under the – in our view – realistic assumptions that competition in health insurance markets is imperfect and healthy consumers are more price elastic than high-risk consumers. If the aim of policy makers is to reduce high-risk consumers' costs in a market with adverse selection, risk adjusters should not merely focus on eradicating all incentives for risk selection. Rather, if they get risk adjustment slightly wrong, they should err on the side of not doing enough rather than doing too much, given the effects on competition. In extreme cases, optimal risk adjustment may even run in the opposite direction. That is, high-risk types are better off if they are taxed to finance a subsidy for the low-risk types. Therefore, it is important that the sponsor of the insurance plan, whether it is a government or an employer, is clear about its objective for risk adjustment.

Finally, our mechanism of competition leverage may be applied also to other sectors that feature some regulation alongside competition. Crucial ingredients are: (i) competing firms that use second-degree price discrimination to separate consumer groups; (ii) heterogeneous price elasticity across these groups; and (iii) government intervention that can differentially affect those contracts/groups. One other example of a market with the potential of competition leverage is telecommunication, where networks use two-part tariffs to separate light from heavy users, whereas the government intervenes by setting interconnection fees. The structure of the regulated interconnection fees will then again affect the intensity of competition.

Appendix A

This appendix contains proofs of propositions.

Proof of Proposition 1. Using (7b) and (7c), we write

$$u^h = \frac{2t}{\lambda f(\frac{1}{2})}(\mu - \tfrac{1}{2}\lambda) + w - \theta^h + \rho^h \tag{A1a}$$

$$u^l = -\frac{2t}{(1-\lambda)g(\frac{1}{2})}(\mu + \tfrac{1}{2}(1-\lambda)) + w - \theta^l + \rho^l - \frac{(\mu\Delta\theta)^2}{\frac{1}{2}r\sigma^2(1-\lambda)^2}. \tag{A1b}$$

To find the effect of μ on u^h, we add λ times equation (A1a) to $(1-\lambda)$ times equation (A1b). This eliminates the direct effects of $\rho^{h,l}$ and the result can be written as:

$$\begin{aligned} u^h &= 2\mu\left(\frac{t}{f(\frac{1}{2})} - \frac{t}{g(\frac{1}{2})}\right) - \frac{\mu^2(\Delta\theta)^2}{\frac{1}{2}r\sigma^2(1-\lambda)} - (1-\lambda)(u^l - u^h) + w - (\lambda\theta^h + (1-\lambda)\theta^l) \\ &\quad -t\left(\frac{\lambda}{f(\frac{1}{2})} + \frac{1-\lambda}{g(\frac{1}{2})}\right) \end{aligned} \tag{A2}$$

where

$$u^l - u^h = \Delta\theta(1-q^l) = \mu\frac{(\Delta\theta)^2}{\frac{1}{2}r\sigma^2(1-\lambda)}\ , \tag{A3}$$

using equations (2a) and (8).

Using (7d), we can rewrite u^h in terms of $1-q^l$:

$$\begin{aligned} u^h &= -(1-\lambda)\Delta\theta(1-q^l) + w - (\lambda\theta^h + (1-\lambda)\theta^l) + (1-\lambda)(1-q^l)r\sigma^2 ME \\ &\quad -t\left(\frac{\lambda}{f(\frac{1}{2})} + \frac{1-\lambda}{g(\frac{1}{2})}\right) - (1-\lambda)r\sigma^2(1-q^l)^2 \end{aligned} \tag{A4}$$

Differentiating u^h with respect to $1-q^l$ yields equation (11a). Writing $u^l = u^h + (u^l - u^h)$, we find

$$\begin{aligned} u^l &= \lambda\Delta\theta(1-q^l) + w - (\lambda\theta^h + (1-\lambda)\theta^l) + (1-\lambda)(1-q^l)r\sigma^2 ME \\ &\quad -t\left(\frac{\lambda}{f(\frac{1}{2})} + \frac{1-\lambda}{g(\frac{1}{2})}\right) - (1-\lambda)r\sigma^2(1-q^l)^2. \end{aligned} \tag{A5}$$

Differentiating this with respect to $1-q^l$ yields (11b).

Aggregate profits can be derived as

$$\Pi = \frac{1}{2}\lambda(w - \theta^h + \rho^h - u^h) + \frac{1}{2}(1-\lambda)(w - \theta^l + \rho^l - \tfrac{1}{2}r\sigma^2(1-q^l)^2 - u^l). \tag{A6}$$

Using the expressions for u^h, u^l as a function of $1-q^l$, we find

$$\Pi = t\left(\frac{\lambda}{f(\frac{1}{2})} + \frac{1-\lambda}{g(\frac{1}{2})}\right) - t(1-\lambda)\left(\frac{1}{f(\frac{1}{2})} - \frac{1}{g(\frac{1}{2})}\right)\frac{r\sigma^2}{\Delta\theta}(1-q^l). \tag{A7}$$

Differentiating this with respect to $1-q^l$ yields equation (11c).

Using equation (10) yields equation (12). *Q.E.D*

Proof of Proposition 2. Assume that there is a choice of $\rho^{h,l}$ that gives an efficient solution: $q^{h,l} = 1$. Note that from the first-order condition (7d), it follows that $q^l = 1$ implies that $\mu = 0$. The incentive compatibility constraints for the high and the low type read

$$IC_h \ : u^h \geq u^l - \Delta\theta(1-q^l)$$

$$IC_l \ : u^l \geq u^h + \Delta\theta(1-q^h).$$

Therefore, efficiency implies $u^h = u^l$. Equations (A1a) and (A1b) with $\mu = 0$ can be written as

$$u^h = w - \theta^h + \rho^h - \frac{t}{f(\frac{1}{2})}$$

$$u^l = w - \theta^l + \rho^l - \frac{t}{g(\frac{1}{2})}.$$

So $u^h = u^l$ leads to

$$\rho^h - \theta^h - \frac{t}{f(\frac{1}{2})} = \rho^l - \theta^l - \frac{t}{g(\frac{1}{2})}. \tag{A8}$$

Conversely, setting the values for ρ such that this holds, evidently solves the first-order conditions plus the IC constraints. Note that there are multiple combinations of ρ^h and ρ^l that yield the first-best outcome. The budget constraint $\lambda\rho^h + (1-\lambda)\rho^l = 0$ pins down a unique pair ρ^l, ρ^h. As $\theta^h > \theta^l$ and $f(\frac{1}{2}) < g(\frac{1}{2})$ (by Assumption (3)), the first best implementing ρ^h and ρ^l satisfies $\rho^h > \rho^l$ and $\rho^h - \theta^h > \rho^l - \theta^l$. *Q.E.D*

Proof of Lemma 1. To derive the relation between $\rho^{h,l}$ and $1-q^l$, subtract the first-order conditions for the utilities, (7c) and (7b), use the one for the low type's coverage q^l (8) to substitute for μ, and the binding incentive compatibility constraint to replace $u^l - u^h = (1-q^l)\Delta\theta$. This results in

$$\rho^h - \rho^l - \frac{t}{f(\frac{1}{2})} + \frac{t}{g(\frac{1}{2})} - \Delta\theta + (1-q^l)\Delta\theta + \tfrac{1}{2}r\sigma^2(1-q^l)^2 + (1-\lambda)\frac{\frac{1}{2}r\sigma^2}{\Delta\theta}\left(\frac{2t}{\lambda f(\frac{1}{2})} + \frac{2t}{(1-\lambda)g(\frac{1}{2})}\right)(1-q^l) = 0.$$

Substituting $\rho^h = -\dfrac{1-\lambda}{\lambda}\rho^l$ (budget balance), it follows that

$$\frac{d\rho^l}{d(1-q^l)} = \lambda\left[\Delta\theta + r\sigma^2(1-q^l) + \frac{\frac{1}{2}r\sigma^2}{\Delta\theta}(1-\lambda)\left(\frac{2t}{\lambda f(\frac{1}{2})} + \frac{2t}{(1-\lambda)g(\frac{1}{2})}\right)\right] > 0. \tag{A9}$$

From this the result follows. *Q.E.D*

Proof of Proposition 3. In view of Lemma 1, a sufficient condition for $\rho^h < 0$ to optimize u^h is that the derivative of u^h to $1-q^l$ is negative at $\rho^h = \rho^l = 0$ (the case without risk adjustment). This derivative is given by Proposition 1 with $\omega = 1, \gamma = 0$, (or equivalently, equation (11a)) to be

$$\frac{du^h}{d(1-q^l)} = (1-\lambda)r\sigma^2\left(-(1-q^l) + ME - \frac{\Delta\theta}{r\sigma^2}\right).$$

Hence, if

$$ME > 1 + \frac{\Delta\theta}{r\sigma^2}. \tag{A10}$$

the optimal ρ^h will be negative. Noting that, by its definition, ME is inversely proportional to $\Delta\theta$, we prove the result. *Q.E.D*

Proof of Proposition 4. By combining the first-order conditions (17a) – (17c) with the "accounting condition" (2a) relating u^l to u^h, as well as budget balance, we find the derivatives of the various utilities in the pooling case, analogously to the expressions in the separating case (11a)–(11c):

$$\frac{du^h}{d(1-q)} = -r\sigma^2(1-q) + r\sigma^2 ME - (1-\lambda)\Delta\theta$$

$$\frac{du^l}{d(1-q)} = -r\sigma^2(1-q^l) + r\sigma^2 ME + \lambda\Delta\theta$$

$$\frac{d\Pi}{d(1-q^l)} = -r\sigma^2 ME.$$

The weighted sum of these expressions gives the required answer in the proposition. From this, the three following observations are immediate. *Q.E.D*

Proof of Proposition 5. For the first part of the proposition, note that

$$(1-\lambda)\frac{dW^{pool}}{d(1-q)} = \frac{dW^{sep}}{d(1-q^l)} + (\omega-\lambda)\lambda\frac{\Delta\theta}{r\sigma^2},$$

at $q = q^l$. Hence, when W^{sep} is maximized, W^{pool} is still increasing in the distortion $1 - q$, if $\omega > \lambda$. This proves the first part.

For the second part, observe that at the first best, $1 - q = 1 - q^l = 0$, $W^{pool} = W^{sep}$. Further, by the above, $\frac{dW^{pool}}{d(1-q)} > \frac{dW^{sep}}{d(1-q^l)}$ for all $1 - q^l \leq (1 - q^l)^{sep}$. Hence, $W^{pool}(1 - q^{sep}) > W^{sep}(1 - q^{sep})$, and *a fortiori* $W^{pool}(1 - q^{pool}) > W^{sep}(1 - q^{sep})$. This proves the second part. *Q.E.D*

Appendix B

This appendix provides a characterization of the Nash equilibrium.

As oligopoly models with price discrimination are not straightforward (see Stole, 2007, for an overview), we need to be careful in characterizing equilibrium outcomes. This section derives sufficient conditions for the Nash equilibrium to be characterized by equations (7a)–(7d).

When insurer I_b chooses u_b^h, u_b^l, we write insurer I_a's optimization problem as

$$\max_{u_a^h, u_a^l, q_a^h, q_a^l} \Pi(u_a^h, u_a^l, q_a^h, q_a^l; u_b^h, u_b^l) \ , \qquad (\hat{P}_{u_b^h, u_b^l})$$

subject to $q_a^h, q_a^l \in [0, 1]$ and the IC constraints for the high type and the low type (2a) and (2b), and where profits are given by

$$\Pi(u_a^h, u_a^l, q_a^h, q_a^l; u_b^h, u_b^l) = F\left(\frac{1}{2} + \frac{u_a^h - u_b^h}{2t}\right)\pi^h(u_a^h, q_a^h) + G\left(\frac{1}{2} + \frac{u_a^l - u_b^l}{2t}\right).\pi^l(u_a^l, q_a^l) \qquad \text{(B1)}$$

Note that I_b's choice of q_b^h, q_b^l does not affect I_a's profits (for given u_b^h, u_b^l). Below, we focus on symmetric Nash equilibria of this optimization problem, defined as follows.

Definition 1. The vector $(u^{h*}, u^{l*}, q^{h*}, q^{l*})$ forms a symmetric Nash equilibrium if $(u^{h*}, u^{l*}, q^{h*}, q^{l*})$ solves $(P_{u^{h*}, u^{l*}})$.

Directly analyzing $(P_{u_b^h, u_b^l})$ is not straightforward for two reasons. First, given I_b's strategy (u_b^h, u_b^l), I_a optimizes over four variables $(u_a^h, u_a^l, q_a^h, q_a^l)$. Second, it is routine to verify that the insurer's optimization problem $(P_{u_b^h, u_b^l})$ is not concave in its four variables. The problem is, however, considerably simplified if we assume that the incentive compatibility constraint for the high type is binding and that the incentive compatibility constraint for the low type is nonbinding. In that case, we can focus on the following relaxed problem.

$$\max_{u_a^h, u_a^l} \Pi\left(u_a^h, u_a^l, 1, 1 - \frac{u_a^l - u_a^h}{\Delta\theta}; u_b^h, u_b^l\right). \qquad (\hat{P}_{u_b^h, u_b^l})$$

This is an optimization problem with only two variables (for given u_b^h, u_b^l). Equations (B2a) and (B2b) below characterize the stationary point u^{h*}, u^{l*} of $(\hat{P}_{u_b^h, u_b^l})$ where $u_b^h = u^{h*}, u_b^l = u^{l*}$:

$$\frac{\lambda}{2t} f\left(\frac{1}{2}\right)\pi^h(u^{h*}, 1) - \frac{\lambda}{2} + \frac{1-\lambda}{2} r\sigma^2 \frac{u^{l*} - u^{h*}}{(\Delta\theta)^2} = 0 \qquad \text{(B2a)}$$

$$\frac{1-\lambda}{2t} g\left(\frac{1}{2}\right)\pi(u^{l*}, q^{l*}) - \frac{1-\lambda}{2} - \frac{1-\lambda}{2} r\sigma^2 \frac{u^{l*} - u^{h*}}{(\Delta\theta)^2} = 0. \qquad \text{(B2b)}$$

Further, we have

$$q^{h*} = 1 \qquad \text{(B3a)}$$

$$q^{l*} = 1 - \frac{u^{l*} - u^{h*}}{\Delta\theta} \in [0, 1]. \qquad \text{(B3b)}$$

We now derive sufficient conditions under which $u^{h*}, u^{l*}, q^{h*}, q^{l*}$, given by equations (B2a)–(B3b), form a symmetric Nash equilibrium as given by Definition 1. We do this in two steps. First, we derive conditions under which u^{h*}, u^{l*} solves $(\hat{P}_{u_b^h, u_b^l})$ where $u_b^h = u^{h*}, u_b^l = u^{l*}$. Second, we derive a condition under which a solution u^{h*}, u^{l*} to $(\hat{P}_{u_b^h, u_b^l})$ is a part of a solution to problem $(P_{u_b^h, u_b^l})$.

A sufficient condition for the first step is that problem $(\hat{P}_{u_b^h, u_b^l})$ is concave.[26] We get the following for the elements of the Hessian:

$$\frac{\partial^2 \Pi_a}{\partial (u_a^h)^2} = \frac{\lambda}{(2t)^2} f'\left(\frac{1}{2} + \frac{u_a^h - u^{h*}}{2t}\right)\pi^h(u_a^h, 1) - \frac{\lambda}{t} f\left(\frac{1}{2} + \frac{u_a^h - u^{h*}}{2t}\right)$$

[26] In fact, quasi-concavity of Π is already sufficient for a stationary point being the global maximum of Π. Olivella and Vera-Hernández (2007) give an example of a Hotelling model where the symmetric equilibrium is characterized by equations (B2a) – (B3b).

$$
\begin{aligned}
&-\lambda F\left(\frac{1}{2}+\frac{u_a^h-u^{h*}}{2t}\right)\frac{r\sigma^2}{(\Delta\theta)^2}\\
\frac{\partial^2\Pi_a}{\partial(u_a^l)^2} &= \frac{1-\lambda}{(2t)^2}g'\left(\frac{1}{2}+\frac{u_a^l-u^{l*}}{2t}\right)\pi^l\left(u_a^l,1-\frac{u_a^l-u_a^h}{\Delta\theta}\right)-\frac{1-\lambda}{t}g\left(\frac{1}{2}+\frac{u_a^l-u^{l*}}{2t}\right)\left(1+r\sigma^2\frac{u_a^l-u_a^h}{(\Delta\theta)^2}\right)\\
&-(1-\lambda)G\left(\frac{1}{2}+\frac{u_a^l-u^{l*}}{2t}\right)\frac{r\sigma^2}{(\Delta\theta)^2}\\
\frac{\partial^2\Pi_a}{\partial u_a^l\partial u_a^h} &= \frac{1-\lambda}{2t}g\left(\frac{1}{2}+\frac{u_a^l-u^{l*}}{2t}\right)r\sigma^2\frac{u_a^l-u_a^h}{\Delta\theta}+(1-\lambda)G\left(\frac{1}{2}+\frac{u_a^l-u^{l*}}{2t}\right)\frac{r\sigma^2}{(\Delta\theta)^2}.
\end{aligned}
$$

First, consider local concavity at $u_a^h = u^{h*}, u_a^l = u^{l*}$. As the density functions f, g are symmetric, we have $f'(\frac{1}{2}) = g'(\frac{1}{2}) = 0$. It then follows that $\frac{\partial^2\Pi_a}{\partial(u_a^h)^2}, \frac{\partial^2\Pi_a}{\partial(u_a^l)^2} < 0$. The determinant of the Hessian in the symmetric equilibrium can be written as

$$
\begin{aligned}
\frac{\partial^2\Pi_a}{\partial(u_a^h)^2}\frac{\partial^2\Pi_a}{\partial(u_a^l)^2}-\left(\frac{\partial^2\Pi_a}{\partial u_a^l\partial u_a^h}\right)^2 &= \frac{\lambda(1-\lambda)}{t^2}f\left(\frac{1}{2}\right)g\left(\frac{1}{2}\right)+\frac{\lambda(1-\lambda)}{2t}f\left(\frac{1}{2}\right)\frac{r\sigma^2}{(\Delta\theta)^2}+\frac{(1-\lambda)^2}{2t}g\left(\frac{1}{2}\right)\frac{r\sigma^2}{(\Delta\theta)^2}\\
&+\frac{1-\lambda}{2t}g\left(\frac{1}{2}\right)r\sigma^2\frac{u^{l*}-u^{h*}}{(\Delta\theta)^2}\left(2\frac{\lambda}{t}f\left(\frac{1}{2}\right)-\frac{1-\lambda}{2t}g\left(\frac{1}{2}\right)r\sigma^2\frac{u^{l*}-u^{h*}}{(\Delta\theta)^2}\right). \quad \text{(B4)}
\end{aligned}
$$

For profits to be locally concave in u_a^h, u_a^l, the determinant of the Hessian must be positive. A sufficient condition for this is

$$
\frac{\lambda}{1-\lambda} > \frac{g\left(\frac{1}{2}\right)}{f\left(\frac{1}{2}\right)}\frac{1}{4\Delta\theta}.
$$

Hence, the share of high-risk types, λ, has to be high enough; a condition that is reminiscent of the condition in Rothschild and Stiglitz (1976) for an equilibrium to exist.

A sufficient condition for equations (B2a)–(B2b) to solve $(\hat{P}_{u_b^h,u_b^l})$ is that Π is globally concave; that is also for $u_a^h \neq u^{h*}, u_a^l \neq u^{l*}$. It follows from the expression of the Hessian above that this leads to an upper-bound on $f'(\frac{1}{2}+\frac{u_a^h-u^{h*}}{2t})$ and $g'(\frac{1}{2}+\frac{u_a^l-u^{l*}}{2t})$ for values of u_a^h, u_a^l where $f' > 0, g' > 0$.

Now we move to the second step: given that u^{h*}, u^{l*} solves $(\hat{P}_{u_b^h,u_b^l})$, when is u^{h*}, u^{l*} together with equations (B3a) and (B3b) a solution to $(P_{u^{h*},u^{l*}})$? That is, when does this form a symmetric Nash equilibrium? To answer this question, we introduce the following notation:

$$
\tilde{u}^l = \arg\max_u (1-\lambda)G\left(\frac{1}{2}+\frac{u-u^{l*}}{2t}\right)\pi^l(u,1)
$$

In words, $\tilde{u}^l$ maximizes profits when the insurer focuses on the l market (where its opponent offers u^{l*} as determined by equations (B2a) and (B2b)) and ignores the h market.

Lemma 2. Let u^{h*}, u^{l*} denote the solution to $(\hat{P}_{u^{h*},u^{l*}})$. Assume that.

$$
\lambda F\left(\frac{1}{2}+\frac{\tilde{u}^l-u^{h*}}{2t}\right)(w-\theta^h+\rho^h-\tilde{u}^l) \geq 0.
$$

Then u^{h*}, u^{l*} together with q^{h*}, q^{l*} as given by equations (B3a)–(B3b) solve $(P_{u^{h*},u^{l*}})$ and hence, form a symmetric Nash equilibrium.

Even if I_a focuses on the l market, θ^h types close to I_a may prefer I_a's l-contract above I_b's h-contract. The assumption makes sure that I_a's l-contract either leads to non negative profits if it is bought by the θ^h type or is not bought by the θ^h type at all.[27] As shown in the proof, this assumption excludes the possibility that a corner solution dominates an interior stationary point in terms of profits.[28]

Proof of lemma 2. In order to simplify notation, we write.

$$
\Pi^*(u^h,u^l,q^h,q^l) = \Pi(u^h,u^l,q^h,q^l;u^{h*},u^{l*}) \quad \text{(B5)}
$$

[27] Note that $\tilde{u}^l$ contract features $q^l = 1$ and hence, type θ^h buying this contract also gets utility $\tilde{u}^l$.

[28] See Olivella and Vera-Hernández (2007) for a numerical analysis of the case with corner solutions.

$$\hat{\Pi}(u^h, u^l) = \Pi^*\left(u^h, u^l, 1, 1 - \frac{u^l - u^h}{\Delta\theta}\right) \tag{B6}$$

Suppose, by contradiction, that the claim is not correct. That is, there exist u^h, u^l, q^h, q^l such that

$$\Pi^*(u^h, u^l, q^h, q^l) > \hat{\Pi}(u^{h*}, u^{l*}). \tag{B7}$$

We consider three cases:

(i) $F(\frac{1}{2} + \frac{u^h - u^{h*}}{2t}) > 0$ and $G(\frac{1}{2} + \frac{u^l - u^{l*}}{2t}) > 0$:

- without loss of generality, $q^h = 1$
  - Suppose not, that is, $q^h < 1$, then increasing q^h increases profits ($\partial \Pi^* / \partial q^h > 0$) and relaxes (IC_l)
- without loss of generality, $q^l = 1 - \frac{u^l - u^h}{\Delta\theta}$; suppose not:
  - if $q^l > 1 - \frac{u^l - u^h}{\Delta\theta}$, (IC_h) is violated
  - if $q^l < 1 - \frac{u^l - u^h}{\Delta\theta}$, (IC_l) is slack and increasing q^l raises profits

but then inequality (B7) contradicts that (u^{h*}, u^{l*}) solves $\hat{P}_{u^{h*}, u^{l*}}$.

(ii) $F(\frac{1}{2} + \frac{u^h - u^{h*}}{2t}) > 0$ and $G(\frac{1}{2} + \frac{u^l - u^{l*}}{2t}) = 0$: we have the following inequalities:

$$\hat{\Pi}(u^{h*}, u^{l*}) \geq \max_u \hat{\Pi}(u, \bar{u}^l) \geq F(\frac{1}{2} + \frac{u^h - u^{h*}}{2t})\pi(u^h, q^h) = \Pi^*(u^h, u^l, q^h, q^l)$$

where $\bar{u}^l = \max\{u | G(\frac{1}{2} + \frac{u - u^{l*}}{2t}) \leq 0\}$. This contradicts equation (B7).

(iii) $F(\frac{1}{2} + \frac{u^h - u^{h*}}{2t}) = 0$ and $G(\frac{1}{2} + \frac{u^l - u^{l*}}{2t}) > 0$: we have the following inequalities:

$$\hat{\Pi}(u^{h*}, u^{l*}) \geq \hat{\Pi}(u^{h*}, \tilde{u}^l) \geq \Pi^*(u^h, u^l, q^h, q^l)$$

where the second inequality follows from the assumption that $F(\frac{1}{2} + \frac{\tilde{u}^l - u^{h*}}{2t})(w - \theta^h + \rho^h - \tilde{u}^l) \geq 0$. Again, we find a contradiction of equation (B7). *Q.E.D*

References

ARMSTRONG, J., PAOLUCCI, F., MCLEOD, H., AND VAN DE VEN, W. "Risk Equalisation in Voluntary Health Insurance Markets: A Three Country Comparison." *Health Policy*, Vol. 98 (2010), pp. 39–49.

ARMSTRONG, M. AND VICKERS, J. "Competitive Price Discrimination." *RAND Journal of Economics*, Vol. 32 (2001), pp. 579–605.

BARDEY, D. AND ROCHET, J.C. "Competition Among Health Plans: A Two-Sided Market Approach. *Journal of Economics & Management Strategy*, Vol. 19 (2010), pp. 435–451.

BARON, D.P. AND MYERSON, R.B. "Regulating a Monopolist with Unknown Costs." *Econometrica*, Vol. 50 (1982), pp. 911–930.

BEAULIEU, N.D. "Quality Information and Consumer Health Plan Choices." *Journal of Health Economics*, Vol. 21 (2002), pp. 43–63.

BIJLSMA, M., BOONE, J., AND ZWART, G. "Community Rating in Health Insurance: Can It Be Optimal?" Mimeo, 2013.

BOSSERT, W. AND FLEURBAEY, M. "Redistribution and Compensation." *Social Choice and Welfare*, Vol. 13 (1996), pp. 343–355.

BUCHMUELLER, T. "Price and the Health Plan Choices of Retirees." *Journal of Health Economics*, Vol. 25 (2006), pp. 81–101.

BUCHMUELLER, T. AND DINARDO, J. "Did Community Rating Induce an Adverse Selection Death Spiral? Evidence from New York, Pennsylvania, and Connecticut." *The American Economic Review*, Vol. 92 (2002), pp. 280–294.

CHERNEW, M.E. AND FRICK, K.D. "The Impact of Managed Care on the Existence of Equilibrium in Health Insurance Markets." *Journal of Health Economics*, Vol. 18 (1999), pp. 573–592.

CUTLER, D., LINCOLN, B., AND ZECKHAUSER, R. "Selection Stories: Understanding Movement across Health Plans." *Journal of Health Economics*, Vol. 29 (2010), pp. 821–838.

CUTLER, D. AND REBER, S. "Paying for Health Insurance: The Tradeoff between Competition and Adverse Selection." *Quarterly Journal of Economics*, Vol. 113 (1998), pp. 433–466.

DAFNY, L. "Are Health Insurance Markets Competitive?" *American Economic Review*, Vol. 100 (2010), pp. 1399–1431.

DEMARZO, P.M., FISHMAN, M.J., AND HAGERTY, K.M. "Self-Regulation and Government Oversight." *The Review of Economic Studies*, Vol. 72 (2005), pp. 687–706.

DOW, W.H., FULTON, B.D., AND BAICKER, K. "Reinsurance for High Health Costs: Benefits, Limitations, and Alternatives." *Forum for Health Economics & Policy*, Vol. 13 (2010). http://econpapers.repec.org/article/bpjfhecpo/default13.htm. Accessed September 12, 2014.

ELLIS, R.P. "Risk Adjustment in Health Care Markets: Concepts and Applications." In M. Lu and E. Jonnson, eds., *Paying for Health Care: New Ideas for a Changing Society*. Weinheim Germany: Wiley-VCH Publishers, 2008.

FAURE-GRIMAUD, A., LAFFONT, J.J., AND MARTIMORT, D. "The Endogenous Transaction Costs of Delegated Auditing." *European Economic Review*, Vol. 43 (1999), pp. 1039–1048.

FLEURBAEY, M. AND SCHOKKAERT, E. "Equity in Health and Health Care." In M.V. Pauly, T.G. McGuire, and P.P. Barros eds., *Handbook of Health Economics*., Vol. 2 of *Handbook of Health Economics*. Amsterdam, The Netherlands: North Holland, 2011.

GLAZER, J. AND MCGUIRE, T.G. "Optimal Risk Adjustment in Markets with Adverse Selection: An Application to Managed Care." *The American Economic Review*, Vol. 90 (2000), pp. 1055–1071.

———. "Setting Health Plan Premiums to Ensure Efficient Quality in Health Care: Minimum Variance Optimal Risk Adjustment." *Journal of Public Economics*, Vol. 84 (2002), pp. 153–173.

HHS. "Patient Protection and Affordable Care Act HHS Notice of Benefit and Payment Parameters for 2014; Proposed Rule." Tech. rep., Department of Health and Human Services, 2012.

HOLMSTRÖM, B. AND MYERSON, R.B. "Efficient and Durable Decision Rules with Incomplete Information." *Econometrica*, Vol. 51 (1983), pp. 1799–1819.

JACK, W. "Controlling Selection Incentives when Health Insurance Contracts are Endogenous." *Journal of Public Economics*, Vol. 80 (2001), pp. 25–48.

———. "Optimal Risk Adjustment with Adverse Selection and Spatial Competition." *Journal of Health Economics*, Vol. 25 (2006), pp. 908–926.

LEDYARD, J.O. AND PALFREY, T.R. "A Characterization of Interim Efficiency with Public Goods." *Econometrica*, Vol. 67 (1999), pp. 435–448.

NUSCHELER, R. AND KNAUS, T. "Health Care Financing Risk Selection in the German Public Health Insurance System." *Health Economics*, Vol. 14 (2005), pp. 1253–1271.

OLIVELLA, P. AND VERA-HERNÁNDEZ, M. "Competition Among Differentiated Health Plans Under Adverse Selection." *Journal of Health Economics*, Vol. 26 (2007), pp. 233–250.

OLSON, J. "Concepts of Equity and Fairness in Health and Health Care." In S. Glied and P. Smith, eds., *Oxford Handbook of Health Economics*. Oxford, UK: Oxford University Press, 2011.

ROTHSCHILD, M. AND STIGLITZ, J. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *The Quarterly Journal of Economics*, Vol. 90 (1976), pp. 629–649.

ROYALTY, A.B. AND SOLOMON, N. "Health Plan Choice: Price Elasticities in a Managed Competition Setting." *The Journal of Human Resources*, Vol. 34 (1999), pp. 1–41.

SASS, H. "The New Triad: Responsibility, Solidarity and Subsidiarity." *Journal of Medicine and Philosophy*, Vol. 20 (1995), pp. 587–594.

SCHMIDT-MOHR, U. AND VILLAS-BOAS, M. "Oligopoly with Asymmetric Information: Differentiation in Credit Markets." *RAND Journal of Economics*, Vol. 30 (1999), pp. 375–396.

SCHOKKAERT, E., DHAENE, G., AND VAN DE VOORDE, C. "Risk Adjustment and the Trade-off between Efficiency and Risk Selection: An Application of the Theory of Fair Compensation." *Health Economics*, Vol. 7 (1998), pp. 465–480.

SCHUT, F.T., GRESS, S., AND WASEM, J. "Consumer Price Sensitivity and Social Health Insurer Choice in Germany and the Netherlands." *International Journal of Health Care Finance and Economics*, Vol. 3 (2004), pp. 117–138.

STAM, P. Testing the Effectiveness of Risk Equalization Models in Health Insurance: A New Method and Its Applications. PhD thesis, Erasmus University Rotterdam, 2007.

STOLE, L. "Nonlinear Pricing and Oligopoly." *Journal of Economics and Management Strategy*, Vol. 4 (1995), pp. 529–562.

———. "Price discrimination and competition." In M. Armstrong and R. Porter, eds., *Handbook of industrial organization*. Amsterdam, The Netherlands: North Holland, 2007.

STROMBOM, B.A., BUCHMUELLER, T.C., AND FELDSTEIN, P.J. "Switching Costs, Price Sensitivity and Health Plan Choice." *Journal of Health Economics*, Vol. 21 (2002), pp. 89–116.

SWARTZ, K. "Reinsuring Risk to Increase Access to Health Insurance." *The American Economic Review*, Vol. 93 (2003), pp. 283–287.

VAN BARNEVELD, E.M., LAMERS, L.M., VAN VLIET, R.C., AND VAN DE VEN, W.P. "Risk Sharing as A Supplement to Imperfect Capitation: A Tradeoff Between Selection and Efficiency." *Journal of Health Economics*, Vol. 20 (2001), pp. 147–168.

VAN DE VEN, W. AND ELLIS, R. "Risk Adjustment in Competitive Health Care Markets." In A.P. Culyer and J.P. Newhouse, eds., *Handbook of Health Economics*. Amsterdam, The Netherlands: Elsevier, 2000.

VAN DE VEN, W. AND SCHUT, F. "Guaranteed Access to Affordable Coverage in Individual Health Insurance Markets." In S. Glied and P. Smith, eds., *Oxford Handbook of Health Economics*. Oxford, UK: Oxford University Press, 2011.

VAN DE VEN, W.P. AND SCHUT, F.T. "Universal Mandatory Health Insurance In The Netherlands: A Model For The United States?" *Health Affairs*, Vol. 27 (2008), pp. 771–781.

VAN VLIET, R. "Free Choice of Health Plans Combined with Risk-Adjusted Capitation Payments: Are Switchers and New Enrolees Good Risks?" *Health Economics*, Vol. 15 (2006), pp. 763–774.