BASIC VERSUS SUPPLEMENTARY HEALTH INSURANCE: THE ROLE OF COST EFFECTIVENESS AND PREVALENCE

by

Jan Boone

October 24, 2014

TILEC Discussion Paper No. 2014-039
CentER Discussion Paper No. 2014-065

ISSN 2213-9532
ISSN 2213-9419
http://ssrn.com/abstract=2516298
Basic versus supplementary health insurance: the role of cost effectiveness and prevalence*

Jan Boone†

October 24, 2014

Abstract

In a model where patients face budget constraints that make some treatments unaffordable, we ask which treatments should be covered by universal basic insurance and which by private voluntary insurance. We argue that both cost effectiveness and prevalence are important if the government wants to maximize the health gain that it gets from its health budget. In particular, basic insurance should cover treatments that are used by people who at the margin buy treatments that are highly cost effective. This is not the same as covering treatments that are themselves highly cost effective.

Keywords: universal basic health insurance, voluntary supplementary insurance, public vs private insurance, access to care, cost effectiveness

JEL classification: I13, I14, D82, H51

---

*Comments by Cedric Argenton and seminar participants in Berlin and Tilburg are much appreciated. Financial support from the Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged.

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

This paper considers a health insurance system where the government offers universal basic health insurance and private parties offer voluntary supplementary insurance. The questions we analyze are: which treatments should be covered by health insurance and how should they be covered; i.e. by basic or supplementary insurance? Should basic insurance cover treatments with the best cost effectiveness scores, the most expensive ones or treatments used by “needy agents” (low income, high risk)?

The motivation for this question is that many countries have such a mixed public/private health care system, while the (health) economics literature does not analyze this question. The Patient Protection and Affordable Care Act (“Obamacare”) is also a move in this direction. Ten essential health benefits categories are defined that need (at least) to be covered by health insurance. It is left to the states to define which treatments within these categories will be covered exactly (McDonough, 2011). People are free to buy more coverage if they want, but any health insurance contract needs to cover at least the basics. So how does one decide which treatments belong to the basics?

We analyze this question in a model where people buy insurance to get access to health care. In particular, agents face a budget constraint for their expenditures on health care. Due to this budget constraint, the agent buys insurance in order to be able to afford a treatment if she needs it. A companion paper, Boone (2014), analyzes basic and supplementary insurance in a standard model with adverse selection and moral hazard. The three main findings in such a model are that (i) basic insurance should cover treatments with the biggest adverse selection problems, (ii) moral hazard plays no role in deciding whether a treatment should be covered in either basic or supplementary insurance; moral hazard does affect the extent to which a treatment is covered, and (iii) cost effectiveness (CE) of a treatment does not affect its priority for basic insurance coverage.

Compared to this benchmark, the access to care model yields three new insights. First, CE scores do play a role in prioritizing which treatments should be insured. Because of limited budgets, money should be spent on treatments with a high health return per dollar spent. Second, treatments with serious moral hazard issues should not be covered at all by (any) insurance. Third, basic insurance should cover treatments that are predominantly used by people who at the margin buy supplementary coverage for the most valuable treatments. The importance of disease (and treatment) prevalence to determine which treatments should be covered by basic insurance appears to be new. We derive sufficient conditions under which basic insurance should target treatments used by people with high risk and low income. Simply covering the most cost effective treatments by basic insurance is likely to waste government resources.

An interesting feature of the access to care model is the definition of moral hazard due to health insurance. The health economics literature tends to define moral hazard as follows. Consider a treatment $k$ that an agent uses if it is covered by health insurance but does not use if she would pay the full costs of $k$ herself (Cutler and Zeckhauser, 2000, pp. 576). Then treatment $k$ is often interpreted as over-consumption induced by moral hazard of insurance. In an access to care model, an agent may not be able to afford treatment $k$ without insurance even though the use of treatment $k$ is socially efficient. Nyman (2003) is one of the first to discuss “efficient moral hazard”. In the model below, we define social efficiency of a treatment...
independently from the question of whether or not it is covered by insurance.

The paper is related to the following strands of literature. First, there is the literature on (health) insurance with adverse selection and moral hazard. Seminal papers are Rothschild and Stiglitz (1976), Pauly (1968) and Zeckhauser (1970). This literature is discussed in Boone (2014). No paper in this literature analyzes the optimal combination of basic and supplementary insurance.

Second, Nyman (1999) and Nyman (2003) stress the access to care motive of health insurance. As these papers point out, there is a difference between health insurance and insurance for property, such as a car. The fact that someone bought the car suggests that she can afford the car and also the repairs that come with it. Then insurance is indeed about reducing consumption risk due to car related expenses. This is different with health insurance. A 50.000 euro cancer treatment may be worth it in terms of utility gained, but a person may not be able to afford it. Nyman does not analyze how the access to care motive affects which treatments should be covered by basic and which by supplementary insurance.

Third, there is a literature on cost effectiveness of treatments. See, for instance, Drummond et al. (2005); Gold et al. (1996); Garber and Sculpher (2011); Meltzer and Smith (2011). The objective of CE analysis is to find the set of treatments that leads to the highest health gain for a given health care budget (e.g. Drummond et al., 2005; Gold et al., 1996). It calculates for each treatment the gain in “health” (usually measured in terms of quality adjusted life years –qaly’s) per unit of the budget spent. Let \( v_k \) denote the health gain (in utility terms) from using treatment \( k \) and \( \delta_k \) the cost of treatment \( k \). We say that treatment \( k \) has a high CE score if \( v_k / \delta_k \) is high.\(^1\)

Roughly speaking, CE analysis proceeds as follows. Rank all treatments in terms of their CE score. Then treatments with the best scores are covered by insurance until the budget is spent. With one exception, the CE literature does not answer the question how treatments should be covered.\(^2\) Although CE analysis can be applied both by governments and private insurers, it seems fair to say that this literature focuses on public insurance. To illustrate, to seriously advice a (profit maximizing) private insurer which treatments to cover, one would have to take selection effects into account. In books like Drummond et al. (2005) and Gold et al. (1996), (adverse) selection is not discussed.

A natural interpretation of this literature is then to say that government sponsored basic insurance should cover the treatments with the highest CE scores. This is intuitive in a model where a treatment is either available through public insurance or not at all.\(^3\) However, we consider a situation where treatments not included in basic can be covered in supplementary insurance.

As mentioned, the underlying idea of CE analysis is to maximize health given the government’s budget constraint. As a consequence “economic evaluations do not usually incorporate the importance of the distribution, of costs and consequences, among different patient

\(^1\)The CE ratio is usually defined as \( \delta_k / v_k \): cost per unit of health gain. We work with the CE score such that a higher score and higher cost effectiveness are “better”.

\(^2\)The only paper in the CE literature –that we are aware of– dealing explicitly with both private and public insurance is Smith (2007, pp. 147). He argues that in a setting with both public and voluntary private insurance, “interventions for the statutory package can be selected solely according to their cost-effectiveness”. In his model, there are no market imperfections in the private market. We augment this literature by allowing for market imperfections in the supplementary segment.

\(^3\)See Hoel (2007) for an extension where a treatment that is not insured can be paid for out-of-pocket.
or population groups, into the analysis ... Rather, an equitable distribution of costs and consequences across socioeconomic or other defined groups in society is viewed as a competing dimension upon which decisions are made, in addition to that of efficient deployment of resources” (Drummond et al., 2005, pp. 47). In an access to care model, maximizing total health, endogenously introduces distributive concerns, as we will see below.

Finally, technically speaking, we analyze a two tiered problem similar to DeMarzo et al. (2005) and Faure-Grimaud et al. (1999). In our case, the government sets parameters for basic insurance which then affect the equilibrium interaction between insurers and consumers on the private market. The health insurance context is similar to Bijlsma et al. (forthcoming), but they consider the effects of risk adjustment.

This paper is organized as follows. The next section discusses health insurance coverage in mixed public-private systems as it is organized around the world. Then we introduce an access to care model. Section 4 introduces the supplementary insurance market. Section 5 characterizes optimal government policy for a government that wants to maximize total health. Here we derive that prevalence is an important factor for basic insurance. Given the government’s optimal policy, section 6 derives which treatments are covered by supplementary insurance. Section 7 derives sufficient conditions under which government uses basic insurance to target “needy” agents. But we also give an example where the government should target resources at low risk types with high income. Section 9 discusses the policy implications of our analysis. The proofs of results are in the appendix.

2. Health insurance coverage in practice

This section gives a brief overview of how public and private insurance are combined in practice. In most OECD countries there exists a combination of public and private health insurance (Colombo and Tapay, 2004, pp. 14). The idea often is that the public system addresses imperfections in the private health insurance market (Blomqvist, 2011; Zweifel, 2011). Roughly speaking, there are three ways in which public and private systems can be combined. First, private insurance may substitute for public insurance. That is, people are either covered privately or publicly. This is the case in Australia, Ireland, Spain and used to be the case in the Netherlands before 2006 (Colombo and Tapay, 2004, pp. 14). Second, private insurance is bought in addition to public insurance to get faster access (i.e. shorter waiting lists), have a broader choice of providers and treatments for a given condition. Examples include Austria, Denmark and Finland (Mossialos and Thomson, 2004, pp. 38/9). Third, private insurance is bought to cover treatment for conditions that are not covered by public insurance. One can think of dental care, physiotherapy and prescription glasses that are not covered by the public insurance system. But also, the public system may not fully cover the costs of a treatment and people can insure the public co-payment on the private market. Examples here include the Netherlands (after 2006), France and Luxembourg (Mossialos and Thomson, 2004, pp. 39-41).

4There are papers trying to address distributive concerns by introducing equity weights into the planner’s objective function (see Meltzer and Smith, 2011; Garber and Sculpher, 2011, for a discussion). For instance, Smith (2005) derives how the optimal co-payment (or user charge) in a purely public system falls with a treatment’s CE score. If the planner’s objective function features equity weights –e.g. in favour of low income people– then treatments used by people with high weights feature lower co-payments as well. As this moves away from the simple idea of maximizing total health, it is hard to implement in practice.
As mentioned, Obamacare is a move in this direction.

Real world systems can have elements from the second and third categories above; e.g. private insurance providing coverage for treatments that are not covered by basic insurance and offering more choice in terms of health care providers. We focus in this paper on the third category (only).\(^5\)

In particular, we consider the case with universal public coverage for a basic insurance package ("essential health benefits" in Obamacare, McDonough, 2011). There is a fixed budget to finance the public system; hence not all treatments can be covered. We assume that there is only one treatment for each condition and we do not model providers. The question is: which conditions should be covered by public insurance and which treatments can be left to the private market? We denote insurance offered by the private voluntary market: supplementary insurance.\(^6\)

### 3. Access to care

This section introduces a simple model that captures that people can experience financial problems to access health care. In particular, we model that there exist treatments that the agent cannot afford if her insurance does not cover these treatments. Further, people may not be able to afford health insurance that covers all available treatments. These access issues have been stressed both in the popular press (Cohn, 2007) and in academic journals (Nyman, 1999; Schoen et al., 2008, 2010). These papers document that without insurance (mainly low income) people forgo treatments that they actually need, because of cost considerations.

Adverse selection models in the tradition of Rothschild and Stiglitz (1976) (RS from here onwards) ignore budget constraints on the side of insured. Nyman (1999) and Nyman (2003) are early papers pointing out that the access motive of insurance is important. We introduce a simpler model than used by Nyman, as we want to characterize which treatments are covered by basic and which by supplementary insurance.

#### 3.1. utility

We model access to care in the following way. An agent has a budget \(\beta\) that she can (at max.) spend on health care. That is, this budget is used to spend on insurance premium, co-payments and buying uninsured treatments. This is modelled in a reduced form way in the sense that we do not derive the budget from a utility function where the consumer decides on allocating income to housing, transport, food, health care etc. We take the budget \(\beta\) for health care expenditure as exogenous.\(^7\) One way to think about this is that people face a credit market constraint. They cannot borrow money to finance a treatment (investment in health). To

\(^{5}\)We ignore private insurance as offering more choice than public insurance, where choice refers to horizontal differentiation. For the same condition, some people prefer treatment \(a\), others \(b\) (some prefer to be treated by provider 1 others by provider 2). Papers modelling this type of choice (and ways in which the insurer can steer patients) include Chernew et al. (2000) and Chernew et al. (2007).

\(^{6}\)Sometimes this is called complementary insurance (see Mossialos and Thomson, 2004, pp. 16).

\(^{7}\)One simple way to get such a budget \(\beta\) for health care is to assume a Cobb Douglas utility function over categories of spending like food, transport, housing and health. Then expenditure on each spending category is a fixed fraction of the agent’s income.
illustrate, a bank may be reluctant to lend money for a liver transplant as the operation itself can signal reduced earning capacity.

We assume that agents are risk neutral for two reasons. First, this allows us to focus on access to care, instead of risk aversion, as a motive for insurance. Boone (2014) analyzes the case of risk averse insured. Second, with utility linear in consumption, a planner who maximizes total welfare, focuses on efficiency (maximizing health) and not on redistribution. As discussed in the introduction, this is in line with the CE literature. However, as we will see below, in an access to care model distributional considerations come to the fore endogenously.

A fraction \( F[1 - F] \) of the population has a low [high] budget \( \beta^l[\beta^h > \beta^l] \) available to spend on health care. We think of people with the high [low] budget \( \beta^h[\beta^l] \) as having high [low] income. Within income class \( i \in \{l, h\} \), a fraction \( \phi^i[1 - \phi^i] \) is low [high] risk type. For each treatment \( k \) the high risk type has higher probability \( \theta^h_k \) of needing the treatment than the low risk type \( \theta^l_k \leq \theta^h_k \).

Further, we know that health and income are positively correlated. Hence a natural assumption is \( \phi^l \leq \phi^h \) (although we will not explicitly use this). Figure 1 gives an overview of the relevant parameters.

---

\(^8\)This implies that types are consistently ranked. See Boone and Schottmüller (2013) for an analysis when this is not the case due to a violation of single crossing.

\(^9\)See, for instance, Frijters et al. (2005), Finkelstein and McGarry (2006), Gravelle and Sutton (2009) and Munkin and Trivedi (2010). Possible explanations for this correlation between income and health include the following. High income people are better educated and hence know the importance of healthy food, exercise etc. Healthy food options tend to be more expensive and therefore better affordable to high income people. Or (with causality running in the other direction) healthy people are more productive and therefore earn higher incomes.

\(^10\)The model is reminiscent of models where agents differ on two dimensions with two types on each dimension: Netzer and Scheuer (2010), Smart (2000) and Wambach (2000). However, in these models more than two IC constraints can be binding leading to the possibility that high and low risk types are pooled in equilibrium. Lemma 2 in the appendix shows that in our model at max. two IC constraints are binding. This keeps our model simple and allows us to analyze basic vs. supplementary insurance.
3.2. treatments

The set of conditions ("illnesses") is denoted \( K = \{1, 2, \ldots, \kappa \} \). For each condition \( k \in K \) there is exactly one treatment; also denoted by \( k \).\(^{11}\)

We assume that conditions are contractible. That is, a physician can determine whether or not a patient suffers from condition \( k \).\(^{12}\) An insurer or the planner can verify this information. Hence, there is no moral hazard on what treatment is needed. However, we do consider moral hazard on whether treatment is needed at all.

We model this as follows. There is a probability \( \psi_k [1 - \psi_k] \) (the same for the \( h \) and \( l \) type, to ease notation) that the patient is in state \( s = 0[1] \). The utility of treatment \( k \) in state \( s \) is denoted by \( v_{sk} > 0 \) and the cost of the treatment is denoted by \( \delta_k \). To ease notation, \( v_{sk} \) does not depend on the patient’s type and \( \delta_k \) does neither depend on the patient’s type nor on the state \( s \).

We assume that treatment is not cost effective in state 0: \( v_{0k} < \delta_k \). But it is in state 1: \( v_{1k} > \delta_k \). Moreover,

\[
v_k \equiv \psi_k v_{0k} + (1 - \psi_k) v_{1k} < \delta_k
\]

In words, if the treatment is used in both states the use of treatment \( k \) destroys social surplus in the sense that the cost exceeds the average value of the health gain. A treatment with high (zero) \( v_{0k} \) features serious (no) moral hazard problems (given that (1) holds). In the terminology of Chandra and Skinner (2012, pp. 663), we consider category II treatments: clear benefits to some patients (group \( 1 - \psi_k \)) but poor average cost effectiveness across all patients.\(^{13}\)

Whereas condition \( k \) is contractible, the severity (state) of the condition is not. We assume that the patient can reveal the state 0 or 1 by truthfully reporting her symptoms to her physician. However, she can also exaggerate her symptoms to qualify for treatment in state 0.\(^{14}\) Hence, a co-payment is needed for the patient to truthfully reveal the state. As utility is linear in consumption, a co-payment

\[
c_k \geq v_{0k}
\]

induces truthful revelation of \( s \) by the patient.

We assume that an agent needs at most one treatment in the period under consideration. This assumption simplifies the exposition considerably. Although the assumption is not innocuous, we view an agent needing two treatments \( k \) and \( k' \) as being in a state where she needs treatment \( m \) which consists of the combination treatments \( k \) and \( k' \). This leads to two advantages. First, we do not impose that the value of treatments can simply be added; we allow for \( v_m \neq v_k + v_{k'} \). Indeed, co-morbidity can affect the value of a treatment. Second, in terms of contracts, we allow for general co-payments. In the state of the world where an agent needs treatments \( k \) and \( k' \), she pays a co-payment equal to \( c_m \) which does not necessarily equal \( c_k + c_{k'} \).

\(^{11}\) We focus on “ex post” treatments. For an analysis of ex ante measures (prevention), see Ellis and Manning (2007).

\(^{12}\) Alternatively, treatment \( k \) for a patient suffering from \( k' \neq k \) yields zero utility.

\(^{13}\) The analysis below can be extended to deal with treatments \( v_k \geq \delta_k \), but it makes the notation rather cumbersome.

\(^{14}\) See Schottmüller (2013) for a formal analysis of this situation as a cheap talk game.
3.3. basic insurance

The cost of treatment $k$ is denoted $\delta_k$. By covering treatment $k$ by basic insurance, the government subsidizes this treatment. The cost of $k$ for the market is denoted by $\gamma_k$. We assume that $\gamma_k \leq \delta_k$: the government does not tax treatments. Each time the treatment is used, the government chips in $\delta_k - \gamma_k$. We say that $k$ is not covered by basic insurance if $\gamma_k = \delta_k$. As basic insurance is universal, the cost reduction applies to all types. The reduced cost shows up in the insurance premium (if $k$ is also covered by supplementary insurance) or in the price of uninsured treatment (if $k$ is not covered by supplementary insurance).

Given $\gamma_k$ set by the government, private insurers can offer coverage of $k$ at co-payment $c_k$. If $c_k = \gamma_k$, treatment $k$ is not covered by supplementary insurance; if $c_k < \gamma_k$ supplementary insurance covers the difference $\gamma_k - c_k$.

As there are four different types $ij$ (with income $\beta^i$ and risk class $\theta^j$, $i, j \in \{l, h\}$), insurers can offer four different contracts indexed by $ij$. Hence, we write $c_{ij}^k$ for the co-payment on treatment $k$ for type $ij$.

Consider an agent $ij$ who sets aside an amount of money $C_{ij}$ (to be determined below) to pay either for co-payments or for uninsured treatments. Given $C_{ij}$, $c_{ij}^k$, we can distinguish three sets of treatments to which the agent has access:

$$
K_{e}^{ij} = \{ k \in K | v_{0k} \leq c_{ij}^k \leq C_{ij} < \gamma_k \}
$$
$$
K_{i}^{ij} = \{ k \in K | v_{0k} > c_{ij}^k \leq C_{ij} < \gamma_k \}
$$
$$
K_{n}^{ij} = \{ k \in K | \gamma_k \leq C_{ij} \}
$$

(3)

The first set of treatments can only be used by the agent if she buys insurance for these treatments as $\gamma_k > C_{ij}$. That is, she cannot pay for it out-of-pocket. If she buys insurance for these treatments, they are used efficiently. Because $c_{ij}^k \geq v_{0k}$, the treatments are only used in state 1. The second set of treatments, if used at all (i.e. if covered by insurance), will be used inefficiently. The co-payment is below $v_{0k}$, hence the treatment is used in both states 0 and 1. Thus treatments in this set reduce total welfare when used (see equation (1)). The final set consists of treatments for which the agent does not need to buy insurance. These treatments can be paid out of $C_{ij}$. As the agent here is risk neutral (she buys insurance to get access to care) and –by construction– needs at max. one treatment, she will not buy insurance for treatments in the set $K_{n}^{ij}$.

To ease the exposition, we assume $\gamma_k \geq v_{0k}$: the government does not induce inefficient health care consumption by making it so cheap that it will be used in both states 0 and 1. To illustrate, allowing for $\gamma_k < v_{0k}$, would lead us to introduce two sets of uninsured treatments ($K_{ne}^{ij}, K_{ni}^{ij}$): one set where uninsured treatments are used efficiently and one where treatments are used inefficiently. Given that most governments worry nowadays about efficient health care utilization, this seems the relevant case to consider.

Treatments with $c_{ij}^k > C_{ij}$ are unavailable for agent $ij$ as she cannot afford the co-payment at the moment that she needs the treatment. This is line with evidence cited by Nyman (2003), Pauly (2008, pp. 116) and Schokkaert and van de Voorde (2011, pp. 334) where high co-payments cause some people to forgo valuable treatments. In the introduction, we discussed a definition of moral hazard that is often used in the health economics literature: if treatment $k$ is used when $k$ is insured, but not without insurance, this use is due to moral hazard (usually
interpreted as over-consumption). In a model with budget constraints, this definition is misleading. Treatments in the set $K_{eij}$ are efficient to use, but without insurance agent $ij$ would not buy them because she cannot afford them.

4. Insurance market

For given $\gamma_k$ set by the government, this section describes what supplementary insurance is offered by insurers. We focus on the second degree price discrimination case: insurers know that there are different types of insured, but they cannot charge different types different prices for the same contract. That is, they cannot risk rate. This can be the case either because they do not observe a customer’s type at the moment that insurance is bought or because the government does not allow them to risk rate (community rating requirement). We think of risk rating (or third degree price discrimination) as a special case where the Lagrange multiplier on the incentive compatibility constraint equals zero.

Further, we assume that the supplementary insurance market is perfectly competitive. Hence, we focus on contracts that are priced in an actuarially fair way (as long as this satisfies incentive compatibility, which it does in our model).

4.1. budget constraints

Given $\gamma_k$ set by the government, if agent $ij$ can buy actuarially fair insurance, her optimal contract is determined by choosing $\rho_{ij}^k, c_{ij}^k$ and $C_{ij}$ which solve:

$$V_{ij} = \max_{C_{ij}, c_{ij}, \rho_{ij}^k} C_{ij} + \sum_{k \in K_{ij}^e} \theta_k^i (1 - \psi_k) \rho_{ij}^k (v_{1k} - c_{ij}^k) + \sum_{k \in K_{ij}^l} \theta_k^i \rho_{ij}^k (v_k - c_{ij}^k) + \sum_{k \in K_{ij}^n} \theta_k^i (1 - \psi_k) (v_{1k} - \gamma_k) - \lambda_{ij} \left( \sum_{k \in K_{ij}^e} \theta_k^i (1 - \psi_k) \rho_{ij}^k (v_{1k} - \gamma_k) + \sum_{k \in K_{ij}^l} \theta_k^i \rho_{ij}^k (\gamma_k - c_{ij}^k) + \beta_i - C_{ij} \right)$$

(4)

where $\rho_{ij}^k \in [0, 1]$ denotes the fraction of supplementary insurance bought by the agent and $\lambda_{ij}$ denotes the Lagrange multiplier on agent $ij$’s budget constraint. As we will see below, for all treatments except one, we have $\rho_{ij}^k \in \{0, 1\}$. The last treatment that the agent covers by insurance can feature $\rho_k \in (0, 1)$ as the agent hits her budget constraint.\(^{15}\)

The first term of $V_{ij}$ is money set aside $C_{ij}$; money not spent on co-payments or uninsured treatments can be spent on other goods. We assume that utility is linear in consumption and normalize the marginal utility of income to 1. In equation (4), there is no $\rho_{ij}^k$ term for treatments $k \in K_{ij}^n$. These treatments are not insured and bought when needed. Treatments $k \in K_{ij}^n$ are

\(^{15}\)In other words, insurance is random in this case: if the agent needs treatment $k$, the insurer randomizes (in a verifiable way) to determine whether the treatment can be reimbursed. In the model, the agent is risk neutral. Hence she does not mind this randomization. We do not suggest lottery-insurance as a serious real world possibility. Indeed, we know that people do not like probabilistic insurance (Wakker et al., 1997). Here we use it to resolve an integer problem as it is not optimal for the insurer to raise the co-payment on this treatment above $C_{ij}$ (which would make the treatment unavailable for the agent).

\(^{16}\)For the $\theta^i$ type there can be two such treatments with $\rho_k \in (0, 1)$ in case of second degree price discrimination: one when she hits her budget constraint and one for the incentive compatibility constraint.
only needed in state 1 and hence the probability that \( \theta^j \) agent needs such a treatment \( k \) equals \( \theta^j_k (1 - \psi_k) \). If the treatment is used, it gives utility \( v_{1k} \) at cost \( \gamma_k \) for the patient. Finally, treatments in \( K_n^j \) do not appear explicitly in the agent’s budget constraint as it is paid out of \( C^j \) in case it is needed.

If \( \rho^j_k = 0 \) for \( k \in K_n^j, K^j_{ij} \), the agent has no access to the treatment because \( \gamma_k > C_{ij} \); treatment \( k \) is too expensive to pay out-of-pocket. For \( \rho^j_k > 0 \), agent buys treatment \( k \in K_e^j \) \( [k \in K_n^j] \) in state 1 \([0 \text{ and } 1]\). Hence, the probability that the treatment is needed equals \( \theta^j_k (1 - \psi_k) [\theta^j_k] \). Utility of the treatment when used is given by \( v_{1k} [v_k] \). The cost of having this treatment covered equals \( \gamma_k \) which is partly paid as co-payment \( (c^j_k) \) and partly \( (\gamma_k - c^j_k) \) as insurance premium. The latter expression appears explicitly in the agent’s budget constraint.

If a treatment is completely covered by basic insurance for \( ij \) \( (\gamma_k = c^j_k) \), it is covered by \( ij \)'s supplementary insurance \( (\rho^j_k = 1) \) without further costs.

The choice of the amount \( C_{ij} \) set aside to finance co-payments affects the budget constraint and the sets \( K_{ij} \) in equation (3). By increasing \( C_{ij} \) less income is available to buy insurance coverage \( \rho_k \), but a larger set of efficient treatments \( K_e^j \cup K_n^j \) becomes available. This is the trade off that the agent faces when deciding on \( C_{ij} \).

Once an agent has decided to put \( C_{ij} \) aside for co-payments and uninsured treatments, it is optimal for the agent to have co-payment \( c^j_k = C_{ij} \) for each treatment \( k \) with \( \gamma_k \geq C_{ij} \) for which she buys coverage. Given that \( C_{ij} \) is available for co-payments, the insurance premium can be reduced by having this co-payment for each treatment. In a standard insurance model, it is optimal to let the co-payment vary per treatment depending on the extent of moral hazard \( v_{0k} \) (Boone, 2014). But with an access to care model as we have here, it is better to determine the highest co-payment that one can afford and then set this co-payment for each treatment for which one buys insurance. As \( c^j_k = C_{ij} \) for each \( k \), we now write \( c^j = C_{ij} \) for the co-payment. By construction, an agent only needs at max. one treatment per period. Recall that this treatment can consist of a number of “sub-treatments”. In this sense, one can think of \( c^j \) as a deductible.

In a standard insurance model with moral hazard and adverse selection, budget constraints do not play a role. People buy insurance for all treatments where value exceeds costs. Less than full insurance happens only due to moral hazard problems or due to adverse selection where low risk types get under-insured in order to separate types (Boone, 2014). Under-insurance because people cannot afford health insurance cannot be captured in the standard framework.

In an access to care model, budget constraints are important. Hence, we think of \( \lambda_{ij} \) as being bigger than 1. Because of the \( C_{ij} \) term, we know that \( \lambda_{ij} \geq 1 \). The case where budget constraints do not bind, is not very interesting. To exclude this case, we assume that

\[
\lambda_{ij} > 1 \quad (5)
\]

for each type \( ij \). Even if both \( ij \) and the government spend their whole budget, there are treatments available that are valuable at the margin.

### 4.2. Incentive Compatibility

As mentioned, we focus on a perfectly competitive supplementary insurance market with second degree price discrimination. We follow the RS definition of the perfect competition equilibrium: (i) each offered contract makes non-negative profits and (ii) given the equilibrium contracts there
is no other contract yielding positive profits.\(^{17}\) Lemma 2 in the appendix shows that in this model with budget constraints, two features of the RS model are preserved: (i) \(\theta^h\) and \(\theta^l\) types are not pooled into one insurance contract\(^{18}\) and (ii) the relevant incentive compatibility (IC) constraint is \(\theta^h\) trying to mimic \(\theta^l\); this happens in each income class. A type in one income class does not want (or is not able) to mimic a type in another income class.

The IC constraint for \(\theta^h\) with income \(\beta^i\) can be written as:\(^{19}\)

\[
V^{ih} \geq \beta^i + \sum_{k \in K^i_l} (1 - \psi_k) \rho^d_k \left( \theta^h_k (v_{1k} - c^i) - \theta^l_k (\gamma_k - c^i) \right) \\
+ \sum_{k \in K^i_l} \rho^d_k \left( \theta^h_k (v_{1k} + (1 - \psi_k)v_{0k} - c^i) - \theta^l_k (\gamma_k - c^i) \right) + \sum_{k \in K^i_n} (1 - \psi_k) \theta^h_k (v_{1k} - \gamma_k)
\]

(6)

The utility of type \(ih\) choosing her own contract equals \(V^{ih}\). If she mimics the \(\theta^l\) type with the same income, her utility equals the sum of the following terms. First, her budget constraint is

\[
\lambda^l \left( \sum_{k \in K^i_l} \theta^l_k (1 - \psi_k) \rho^d_k (\gamma_k - c^i) + \sum_{k \in K^i_l} \theta^l_k \rho^d_k (\gamma_k - c^i) - (\beta^i - c^i) \right) \\
- \lambda^l \left( \sum_{k \in K^i_l} (1 - \psi_k) \rho^d_k (\theta^h_k (v_{1k} - c^i) - \theta^l_k (\gamma_k - c^i)) \\
+ \sum_{k \in K^i_n} \rho^d_k \left( \theta^h_k (v_{1k} + (1 - \psi_k)v_{0k} - c^i) - \theta^l_k (\gamma_k - c^i) \right) + \sum_{k \in K^i_n} (1 - \psi_k) \theta^h_k (v_{1k} - \gamma_k) - (V^{ih} - \beta^i) \right)
\]

where \(\nu^d\) is the Lagrange multiplier on (IC).\(^{20}\)

\(^{17}\)In RS an equilibrium may not exist if the fraction of \(\theta^l\)-types is large. This paper does not focus on existence of equilibrium. Implicitly, we assume that the fraction of \(\theta^l\)-types is small enough that an equilibrium exists.

\(^{18}\)Assuming they are different given the treatments they can afford with their budgets: see lemma 2.

\(^{19}\)Substituting \(\beta^i\) from the budget constraint gives a right hand side similar to (4). But here it is convenient to start the right hand side with \(\beta^i\) instead of \(C^d\).

\(^{20}\)Budget \(\beta^i\) is moved to the last term \(V^{ih} - \beta^i\) for reasons that become clear in the discussion of lemma 1 in section 7.
The timing of the game is as follows. In the first period, government decides which treatments to cover by basic insurance (i.e. it sets $\gamma_k$ for each $k \in K$). Then insurers choose contracts. We focus on second degree price discrimination. However, risk rating is captured by $\nu^{il} = 0$ in equation (6): insurers can discriminate between types in terms of premium without worrying about (IC). From the contracts that they can afford, consumers buy the insurance contract which maximizes their expected utility. If an insured falls ill, she gets treatment if either (i) she has supplementary insurance for this treatment (and pays the co-payment) or (ii) she can afford to buy it without supplementary insurance.

At first sight, one would expect that backward induction is needed to find the government’s optimal policy. However, because supplementary insurance is chosen to maximize agents’ utility, it turns out that we can use an envelop argument (see appendix B for details). This allows us to characterize the government’s optimal basic insurance package before deriving the contracts offered on the supplementary market. The next section derives the optimal $\gamma_k$ for $k \in K$. Section 6 derives which treatments are covered by supplementary insurance for each type $ij$ and section 8 derives the optimal co-payments for each contract.

5. Optimal government policy

The government chooses $\gamma_k \in [\psi_k, \delta_k]$ to maximize total welfare, defined as

$$W = F(\phi^l V^{il} + (1 - \phi^l) V^{ih}) + (1 - F)(\phi^h V^{il} + (1 - \phi^h) V^{ih})$$

where $V^{th}$ [$V^{il}$] is defined in equation (4) [(6)]. We follow the CE literature where the government maximizes total health without risk and/or income solidarity considerations.

The planner has a budget $B$ that can be spent on basic insurance. The budget constraint can be written as\(^{21}\)

$$F\phi^l \left( \sum_{k \in K^{il}} \theta^l_k (1 - \psi_k) \rho^{il}_k (\delta_k - \gamma_k) + \sum_{k \in K^{il}} \theta^l_k \rho^{il}_k (\delta_k - \gamma_k) + \sum_{k \in K^{il}} \theta^l_k (1 - \psi_k)(\delta_k - \gamma_k) \right) +$$

$$F(1 - \phi^l) \left( \sum_{k \in K^{ih}} \theta^h_k (1 - \psi_k) \rho^{ih}_k (\delta_k - \gamma_k) + \sum_{k \in K^{ih}} \theta^h_k \rho^{ih}_k (\delta_k - \gamma_k) + \sum_{k \in K^{ih}} \theta^h_k (1 - \psi_k)(\delta_k - \gamma_k) \right) +$$

$$(1 - F)\phi^h \left( \sum_{k \in K^{lh}} \theta^l_k (1 - \psi_k) \rho^{lh}_k (\delta_k - \gamma_k) + \sum_{k \in K^{lh}} \theta^l_k \rho^{lh}_k (\delta_k - \gamma_k) + \sum_{k \in K^{lh}} \theta^l_k (1 - \psi_k)(\delta_k - \gamma_k) \right) +$$

$$(1 - F)(1 - \phi^h) \left( \sum_{k \in K^{lh}} \theta^h_k (1 - \psi_k) \rho^{lh}_k (\delta_k - \gamma_k) + \sum_{k \in K^{lh}} \theta^h_k \rho^{lh}_k (\delta_k - \gamma_k) + \sum_{k \in K^{lh}} \theta^h_k (1 - \psi_k)(\delta_k - \gamma_k) \right) \leq B$$

(GBC)

For each type $ij$, the equation gives the probability that this type uses treatment $k$ and hence the government needs to contribute $\delta_k - \gamma_k$ for treatment of this type. If either treatment $k$ is

\(^{21}\)In a system where mandatory basic insurance is not financed by the government but sold by private insurers, the relevant constraint would be that low income people can afford to buy the basic insurance package. This leads to a similar analysis.
not covered by basic insurance \((\gamma_k = \delta_k)\) or it is not used by anyone \((k \notin K_n^{ij}\) and \(\rho_{k}^{ij} = 0\) for each \(ij\), the government does not need to contribute anything for this treatment.

In order to characterize the optimal values of \(\gamma_k\), we define two indicator functions:

\[
I_k^{ij} = \begin{cases} 
1 - \psi_k & \text{if } k \in K_e^{ij} \\
1 & \text{if } k \in K_i^{ij} \\
0 & \text{otherwise}
\end{cases} \tag{8}
\]

\[
I_{\gamma_k \leq c^{ij}} = \begin{cases} 
1 & \text{if } k \in K_n^{ij} \\
0 & \text{otherwise}
\end{cases} \tag{9}
\]

The function \(I_k^{ij}\) gives the fraction of times that an insured treatment is used (conditional on the condition occurring). Treatment \(k \in K_e^{ij}\) is only used in state 1 and \(k \in K_i^{ij}\) is used in both states 0 and 1. \(I_{\gamma_k \leq c^{ij}}\) indicates that a treatment is bought without insurance.

**Proposition 1** Rank treatments \(k\) on the basis of

\[
\sum_{i \in \{l,h\}} \left\{ \sum_{j \in \{l,h\}} \pi_k^{ij} \left[ (1 - I_{\gamma_k \leq c^{ij}})\lambda^{ij} + I_{\gamma_k \leq c^{ij}} \right] + F^i \phi^j \lambda^{ij} \left( \theta_k^l \rho_k^l - \theta_k^h \rho_k^h \right) \right\} \tag{10}
\]

\[
\sum_{i' \in \{l,h\}} \sum_{j' \in \{l,h\}} F^{i'} \phi^{j'} I_k^{ij'} \left( \rho_k^{ij'} (1 - \psi_k I_{\gamma_k \leq c^{ij'}}) + (1 - \psi_k) I_{\gamma_k \leq c^{ij'}} \right)
\]

with \(\gamma_k \in [v_k, \delta_k]\) taking on one of the values \(\{v_k, c^{il}, c^{ih}, c^{hl}, c^{hh}\}\), where

\[
\pi_k^{ij} = \frac{F^i \phi^j \lambda^{ij} \left( \rho_k^l + (1 - \psi_k) I_{\gamma_k \leq c^{ij}} \right)}{\sum_{i' \in \{l,h\}} \sum_{j' \in \{l,h\}} F^{i'} \phi^{j'} I_k^{ij'} \left( \rho_k^{ij'} + (1 - \psi_k) I_{\gamma_k \leq c^{ij'}} \right)} \tag{11}
\]

with \(F^l = F, F^h = 1 - F, \phi^l = \phi, \phi^h = 1 - \phi\). Basic insurance covers highest ranking treatments \(k\) until the budget \(B\) runs out. Treatments that are not covered have \(\gamma_k = \delta_k\).

By subsidizing treatments through basic insurance, the government frees up agents’ resources. These freed up resources can be spent by the agent on buying either additional insurance or other goods. Further, basic insurance helps to reduce adverse selection by relaxing IC constraints. The proposition says that treatments \(k\) should be covered by basic insurance where the returns to agents’ freed resources and relaxed IC constraints are highest.

To understand the expression in (10), first consider the case where IC constraints are not binding \((\nu^{ij} = 0)\). This is the case with risk rating, but can also happen with community rating (see example 1 below).

Note that the sum of \(\pi_k^{ij}\) over \(ij\) adds up to 1. Hence, we rank on the basis of a weighted average. Consider a type \(ij\) and assume that \(\gamma_k > c^{ij}\). If \(ij\) has no insurance for \(k\) \((\rho_{k}^{ij} = 0)\), she has no access to this treatment and a small reduction in \(\gamma_k\) has no effect on \(ij\)’s utility: \(\pi_k^{ij} = 0\). If \(k\) is insured \((\rho_{k}^{ij} > 0)\) then \(\pi_k^{ij} > 0\) and reducing \(\gamma_k\) leads to a reduction in \(ij\)’s insurance premium. The money saved on the premium can be used to buy insurance for other treatments. At the margin, this additional money has return equal to \(\lambda^{ij}\). This follows from equation (4): \(dV^{ij}/d\beta^i = \lambda^{ij}\). Finally, if \(k\) is bought without insurance (i.e. paid out of \(C^{ij} = c^{ij}\), reducing
illustrates, these are not necessarily treatments with a high CE score. Hence, the idea in 
(Drummond et al., 2005; Gold et al., 1996) is not correct in a context where supplementary
the CE literature that the government should cover treatments with the highest CE score
insurance is available.

features: \( \rho \)

a higher return. Hence (compared to other types) and for which
in this cost. In other words, we take the weighted average of the return to a reduction in
ij
is achieved by covering treatments
k
for diseases that are particularly prevalent under
mn
that give high weight to type
k
mn
constitute an equilibrium.

Equation (10) gives the return for the planner of a small change in \( \gamma_k \).

Equation (10) shows the importance of prevalence of a disease. Let \( \lambda^{mn} \) denote the highest \( \gamma^i \) among all types. Note that \( \lambda^i \) does not vary with \( k \), it is the weight \( \pi^i_k \) that varies with \( k \). Hence, basic insurance should cover treatments \( k \) that give high weight to type \( mn \). This is achieved by covering treatments \( k \) for diseases that are particularly prevalent under \( mn \) (compared to other types) and for which \( mn \) buys insurance \( \rho_k^{mn} > 0 \). As the next example illustrates, these are not necessarily treatments with a high CE score. Hence, the idea in the CE literature that the government should cover treatments with the highest CE score (Drummond et al., 2005; Gold et al., 1996) is not correct in a context where supplementary insurance is available.

Example 1 This example illustrates two points made above: (i) IC constraints do not necessarily bind with second degree price discrimination and (ii) basic insurance should not cover treatments with the highest CE score. To make these points, we assume that there are three treatments: \( K = \{1, 2, 3\} \). These treatments do not suffer from moral hazard: \( v_{0k} = \psi_k = 0 \) for each \( k \in K \). Co-payment equals \( c = 0 \) for each type. Table 1 gives the value created by each treatment, its cost and the probability that the treatment is needed for each \( \theta \) type. People with low income have \( \beta_l^i = 1, \phi_l^i = 0.5 \) and with high income: \( \beta_h^i = 2, \phi_h^i = 0.5 \).

First, consider the case where there is no government budget (\( B = 0 \)). Market equilibrium features: \( \rho_{1l}^i = \rho_{1h}^i = 1 \) and \( \rho_{2l}^i = \rho_{2h}^i \). Hence the low income types are pooled on a contract with premium \( \sigma_l^i = \sigma_h^i = \beta_l^i = 1 \) due to the budget constraint there is no adverse selection for \( \beta_l^i \) types. This contract covers the treatment with the highest CE score: \( v_{11}/\delta_1 = 3 \).

\[\begin{array}{c|cccc}
K & v_{1k} & \delta_k & \theta_k^h & \theta_k^l \\
1 & 30 & 10 & \frac{1}{10} & \frac{1}{10} \\
2 & 20 & 10 & \frac{1}{5} & 0 \\
3 & 15 & 10 & \frac{1}{6} & \frac{1}{6} \\
\end{array}\]

Table 1: Value, costs and probabilities for treatments \( K \) in example 1.

\( \gamma_k \) only increases consumption on outside goods (weight one: \( I_{\gamma_k \leq c} \) is multiplied by 1). That is, this does not free up money to buy more health insurance.

We take the average of these returns over \( ij \) with weight \( \pi^i_j \). The denominator of \( \pi^i_j \) is the cost for the government of reducing \( \gamma_k \) by 1 euro. The numerator of \( \pi^i_j \) is the weight of type \( ij \) in this cost. In other words, we take the weighted average of the return to a reduction in \( \gamma_k \) where type \( ij \)'s weight equals her share in the cost of this reduction in \( \gamma_k \).

If \( c = c \) for all \( ij \), reducing \( \gamma_k \) below \( c \) is not optimal due to assumption (5). This would yield a return equal to 1, while (5) implies that there are treatments that can be covered with a higher return. Hence \( \gamma_k < c \) has a very low ranking in equation (10). If not all \( c \) are the same, \( \gamma_k < c \) for some \( ij \) is possible. However, basic insurance does not cover treatments with low costs: \( \delta_k < \min_{ij} c \) as long as (5) holds. Since both the planner’s objective and the budget constraint are linear in \( \gamma_k \), \( \gamma_k \) can at max. have one of five different values: either \( v_{0k} \) (fully covered by basic insurance) or one of the values \( c \). Equation (10) gives the return for the planner of a small change in \( \gamma_k \).

Equation (10) shows the importance of prevalence of a disease. Let \( \lambda^{mn} \) denote the highest \( \gamma^i \) among all types. Note that \( \lambda^i \) does not vary with \( k \), it is the weight \( \pi^i_k \) that varies with \( k \). Hence, basic insurance should cover treatments \( k \) that give high weight to type \( mn \). This is achieved by covering treatments \( k \) for diseases that are particularly prevalent under \( mn \) (compared to other types) and for which \( mn \) buys insurance \( \rho_k^{mn} > 0 \). As the next example illustrates, these are not necessarily treatments with a high CE score. Hence, the idea in the CE literature that the government should cover treatments with the highest CE score (Drummond et al., 2005; Gold et al., 1996) is not correct in a context where supplementary insurance is available.

Example 1 This example illustrates two points made above: (i) IC constraints do not necessarily bind with second degree price discrimination and (ii) basic insurance should not cover treatments with the highest CE score. To make these points, we assume that there are three treatments: \( K = \{1, 2, 3\} \). These treatments do not suffer from moral hazard: \( v_{0k} = \psi_k = 0 \) for each \( k \in K \). Co-payment equals \( c = 0 \) for each type. Table 1 gives the value created by each treatment, its cost and the probability that the treatment is needed for each \( \theta \) type. People with low income have \( \beta_l^i = 1, \phi_l^i = 0.5 \) and with high income: \( \beta_h^i = 2, \phi_h^i = 0.5 \).

First, consider the case where there is no government budget (\( B = 0 \)). Market equilibrium features: \( \rho_{1l}^i = \rho_{1h}^i = 1 \) and \( \rho_{2l}^i = \rho_{2h}^i = \rho_{3l}^i = \rho_{3h}^i = 0 \). Hence the low income types are pooled on a contract with premium \( \sigma_l^i = \sigma_h^i = \beta_l^i = 1 \) due to the budget constraint there is no adverse selection for \( \beta_l^i \) types. This contract covers the treatment with the highest CE score: \( v_{11}/\delta_1 = 3 \).

\[22\text{The general characterization of supplementary insurance is given in proposition 2 below. In this simple example it is routine to verify that these values of } \rho^i_k \text{ constitute an equilibrium.}\]
For the high income types we find the following: $\rho_{1h}^h = 1, \rho_{2h}^h = 1/2, \rho_{3h}^h = 0$ at a premium equal to $\sigma^h = \beta^h = 2; \rho_{1l}^l = 1, \rho_{2l}^l = 0, \rho_{3l}^l = 1/2$ at a premium equal to $\sigma^l = \beta^l = 2$. These contracts are incentive compatible.

Now suppose that the government has a small budget $B > 0$; which treatments should be covered by basic insurance? Based on the idea from the CE literature that the government should cover treatments with highest CE scores, one option is to cover (part of) treatment 1. This would free up money for all types to buy additional supplementary insurance. The $\theta^h$ types (both with low and high income) buy more coverage of treatment 2, the $\theta^l$ types buy more coverage of treatment 3. Hence, the health gain per unit of budget $B$ spent is a weighted average of $v_{12}/\delta_2 = 2.0$ and $v_{13}/\delta_3 = 1.5$.

Equation (10) suggests to cover treatment 2 in basic insurance. Marginal utility of income equals 2 for $\theta^h$ and 1.5 for $\theta^l$. Hence, basic insurance should cover the treatment that is predominantly used by $\theta^h$ which is treatment 2. For $B > 0$ small enough, this yields a health gain of 2.0 per unit of the budget $B$ spent. Hence covering treatment 1 by basic insurance is, in fact, suboptimal.

Recall that the planner’s goal is to maximize health (efficiency) without any distributive objective. But in an access to care model, this efficiency objective endogenously gives the planner an incentive to redistribute towards types with high $\lambda$. To maximize the health impact of the budget $B$, basic insurance should cover/subsidize treatments used by types that at the margin buy insurance coverage for treatments with a high CE score. As the example shows this is not the same as basic insurance covering treatments with high CE scores (as they may also be used by types with a low return $\lambda$ at the margin). Section 7 derives a sufficient condition under which high $\lambda$ types (“most deserving”) are, in fact, people with low income and low health (“most needy”).

Going back to proposition 1, the ranking of treatments for basic insurance is based on the sum of two terms in equation (10). The second term shows that basic insurance should cover/subsidize treatments used by types that at the margin buy insurance coverage for treatments with a high CE score. As the example shows this is not the same as basic insurance covering treatments with high CE scores (as they may also be used by types with a low return $\lambda$ at the margin). Section 7 derives a sufficient condition under which high $\lambda$ types (“most deserving”) are, in fact, people with low income and low health (“most needy”).

Two terms determine the extent to which a reduction in $\gamma_k$ reduces the adverse selection inefficiency. First, there is

$$\theta_{ik}^h \gamma_k^h \rho_{kk}^h - \theta_{ik}^l \gamma_k^l \rho_{kl}^l$$

(12)

where $\theta_{ik}^h - \theta_{ik}^l > 0$ tends to move treatment $k$ up in priority. A high value of $\theta_{ik}^h - \theta_{ik}^l$ implies that private coverage of $k$ for $\theta^h$ makes it attractive for $\theta^h$ to mimic $\theta^l$: probability that this type needs treatment is given by $\theta_{ik}^h$ while the insurance premium is determined by $\theta_{ik}^l$. Hence, the incentive to mimic is related to $(\theta_{ik}^h - \theta_{ik}^l)\gamma_k$. This incentive (and the ensuing distortion) is reduced as $\gamma_k$ falls. In a standard model of moral hazard and adverse selection this is the deciding factor in choosing treatments for basic insurance (Boone, 2014). There the effect always works in the direction of covering treatments with highest $\theta_{ik}^h - \theta_{ik}^l$ by basic insurance.

Here, however, because agents are budget constrained, $\theta_{ik}^h - \theta_{ik}^l > 0$ can work in the opposite direction for some treatment $k$. Due to budget constraints, there can be treatments $k$ with $\rho_{kl}^l > \rho_{kl}^h$ making the expression in (12) negative. Then a reduction in $\gamma_k$ makes mimicking more attractive as it gives mimicking $ih$ cheaper access to $k$. $\nu^l > 0$ implies a lower ranking for treatments where (12) is negative.
There is a second reason why reducing $\gamma_k$ may encourage mimicking. This happens if

$$\theta^h_k[I^h_k \rho^h_k (\lambda_{ih} - 1) + (1 - \psi_k)(I_{\gamma_k \leq c_{ih}} - I_{\gamma_k \leq c_{il}})] < 0$$

in equation (10). To see the intuition for this expression, consider the case where $c_{ih} < c_{il}$. Then a treatment with $v_{0k} \leq c_{ih} < \gamma_k < c_{il}$ is bought without insurance by $il$ but not by $ih$. There are two cases to consider. First, $\rho^h_k = 0$: type $ih$ does not have access to $k$ at all. Then a reduction in $\gamma_k$ has no effect on $ih$’s utility (when buying the $ih$-contract) while it raises the payoff of mimicking $il$. Hence, reducing $\gamma_k$ encourages mimicking. Second, $\rho^h_k > 0$. Then reducing $\gamma_k$ reduces $ih$’s insurance premium allowing $ih$ to buy more insurance at the margin. Compared to buying $k$ without insurance, this has an additional effect $\lambda_{ih} - 1 > 0$ on the utility of $ih$ (when buying the $ih$-contract). If the overall effect is still negative, the reduction in $\gamma_k$ encourages mimicking.

Summarizing, basic insurance should cover treatments that (i) are predominantly used by agents with high $\lambda_{ij}$ and (ii) help to reduce the adverse selection inefficiency.

6. Supplementary insurance

As shown in proposition 1, the CE score is not leading when it comes to basic insurance. Here we show that CE scores are important for supplementary insurance. We define the following CE score for a treatment $k$ with co-payment $c_{ij}$ and cost on the private market $\gamma_k$.

$$s_k(\gamma_k, c_{ij}) = \begin{cases} v_{1k} - \gamma_k & \text{if } c_{ij} \geq v_{0k} \\ \gamma_k - c_{ij} & \text{if } c_{ij} < v_{0k} \end{cases}$$

where we take $c_{ij}$ as given; section 8 endogenizes $c_{ij}$. Treatments that are not covered by basic insurance have $\gamma_k = \delta_k$ and equation (1) then implies $s_k < 0$ for $c_{ij} < v_{0k}$.

**Proposition 2** Given $c_{ij}$, supplementary insurance does not cover treatments with $\gamma_k < c_{ij}$.

For $\theta^h$ agents, insurers rank treatments on the basis of their CE score

$$s_k(\gamma_k, c_{ih})$$

(14)

For $\theta^l$ types, insurers rank treatments on the basis of the adjusted CE score

$$s_k(\gamma_k, c_{il}) \left(1 - \nu_{il} \left[1 + \frac{1 + s_k(\gamma_k, c_{il})}{s_k(\gamma_k, c_{il})} \left(\frac{\theta^h_k}{\theta^l_k} - 1\right)\right]\right)$$

(15)

Supplementary insurance for type $ij$ covers the highest ranking treatments until agent $ij$’s budget $\beta^i - c_{ij}$ runs out.

With $\nu_{il} = 0$, both $\theta^l$ and $\theta^h$ types use supplementary insurance to cover treatments with the highest CE scores (14). Assumption (5) rules out the case where enough budget is available to cover all valuable treatments. Hence, agents choose the treatments that give the highest (health) utility per unit of their budget spent. In an insurance model without budget constraints, all valuable treatments can be covered and hence there is no reason to prioritize on the basis of CE scores (Boone, 2014).
Although this is not mentioned in handbooks on CE analysis, the CE score should actually be corrected for the co-payment $c_{ij}$. Hence, the correct CE score is not simply $v/\delta$, but the expression in equation (13). This is a minor adjustment but it does give rise to an interesting inverse-U relation between cost of treatment and coverage by insurance. To see this, fix $c_{ij} \geq v_{0k}$ and $v_{1k} > c_{ij}$ and let $\gamma_k$ increase from $v_{0k}$. Then for $\gamma_k < c_{ij}$, the treatment should not be covered by supplementary insurance: people can pay the treatment out-of-pocket. That is, relatively cheap treatments should not be covered by supplementary insurance at all. Treatments with $\gamma_k > c_{ij}$ close to $c_{ij}$ will certainly be covered by insurance as the ratio in (13) is very high for such treatments. A small investment (in insurance) $\gamma_k - c_{ij}$ leads to a big health gain. Then as $\gamma_k$ keeps rising (for given $v_{1k}$), the CE score deteriorates and $\lambda_k$ drops in priority. For $\gamma_k$ high enough, the treatment will not be covered by insurance. A second observation is that CE scores actually differ between persons with different co-payments $c_{ij}$. Hence there is not an unique CE score for a treatment.

Finally, treatments with severe moral hazard problems ($v_{0k} > c_{ij}$) and $\gamma_k = \delta_k$ are not covered by supplementary insurance. Hence in this model, moral hazard affects the extensive margin of insurance. In a standard model, moral hazard leads to marginally lower insurance by requiring a higher co-payment (Pauly, 1968; Zeckhauser, 1970). That is, moral hazard affects the intensive margin only, not the external margin.

The effect of $\nu^{il} > 0$ on the treatment ranking for supplementary insurance for $\theta^l$ is as follows. For $\nu^{il} > 0$ there is a “correction term” in equation (15). This correction disappears if $\theta^h_k = \theta^l_k$; treatments that are used equally by $\theta^h$ and $\theta^l$ types, do not cause adverse selection issues and hence do not need to be distorted in supplementary insurance. However, if $\theta^h_k > \theta^l_k$, the use of this treatment will be distorted in order to separate the $\theta^l$ from the $\theta^h$ contract. The distortion takes the form of pushing this treatment down the ranking for $\theta^l$. Hence $\theta^l$ may end up without coverage of this treatment even though its CE score (13) is quite high. Treatments $k$ with high $(1 + s_k)/s_k$ (i.e. low $s_k$) get pushed down further in the ranking (for given $\theta^h_k/\theta^l_k$). If $v - \gamma$ is high, $\theta^h$ does not want to miss out on this treatment and a small distortion on this treatment is enough to stop $\theta^h$ from buying the $\theta^l$-contract. If $\gamma_k - c$ is small, the gain in insurance premium of mimicking is small for the $\theta^h$ type (this gain is of the form $(\theta^h_k - \theta^l_k)(\gamma_k - c_{il})$). In either case, $s_k$ is rather high and treatment $k$ does not need to be distorted much to get separation.

7. Redistribution

The following lemma gives a sufficient condition for basic insurance to focus on the “most needy” agents: low income and high risk. That is, basic insurance should cover treatment of conditions that are prevalent under low income/high risk agents, for which they buy insurance.

**Lemma 1** Assume that $C^{ij} = c^{ij} = c$ for all types $ij$. Then $\lambda^h \geq \lambda^h$ and $\lambda^l \geq \lambda^l$ for $i \in \{l, h\}$.

The co-payment $c$ can be the same for all types for two reasons. First, the government mandates co-payment $c$ for supplementary insurance. Second, all types choose the same co-

\footnote{Clearly, ranking on the basis of $(v - \delta)/\delta$ or inverse ranking on the CE ratio $\delta/v$ gives the same results as ranking on the basis of $v/\delta$.}

\footnote{Note that $\nu^{il}$ does not vary with $k$.}
payment (see next section).

The lemma shows that we can rank the marginal utilities of income \( \frac{dV^{ij}}{d\beta^i} = \lambda^{ij} \). Type \( lh \) has the highest marginal utility of income of all four types. Further, within an income class \( i \) the \( \theta^h \) type has higher marginal utility of income than the \( \theta^l \) type. This happens for two reasons. First, an agent with either a higher income or lower probabilities \( \theta_k \) can cover more treatments with supplementary insurance. As the treatments are ranked, this implies that the agent with a higher budget and/or lower risk covers lower ranked treatments at the margin. Thus, the marginal utility of income falls. Second, \( \theta^l \) types use a distorted ranking (15) and hence, at the margin, cover treatments with lower value.

Comparing the marginal utility of income for two \( \theta^l \) types (one with high, the other with low income) is not straightforward due to the interaction of the budget and IC constraints. As \( \beta \) increases (from \( \beta^l \) to \( \beta^h \)) the budget constraint relaxes. Because \( \frac{dV^{ih}}{d\beta^i} \geq 1 \), we have \( \frac{d(V^{ih} - \beta^i)}{d\beta^i} \geq 0 \). Hence also the IC constraint in (6) is relaxed in response to an increase in \( \beta^i \). As both the budget and the IC constraints are relaxed, it is natural to expect that both shadow prices \( (\lambda^{il}, \nu^{il}) \) fall as well.\(^{25}\)

Combining proposition 1 and lemma 1, basic insurance’s focus on treatments \( k \) with high prevalence among high risk types, leads to a focus on treatments where \( \theta^h_k - \theta^l_k \) is high. Basic insurance covering treatments with high \( \theta^h_k - \theta^l_k \) for which \( \theta^h \) types buy insurance, helps to target money at high risk types. This frees up money for these people to buy additional insurance which leads to high health gains at the margin (high \( \lambda \)).

This observation is significant in the light of the CE literature. As discussed in the introduction, the CE literature tends to view distribution as orthogonal to maximizing health gain per unit of the government budget spent. In an access to care model, by maximizing total health (7), the planner endogenously takes distribution into account. Government budget should be targeted at agents with highest \( \lambda \).

The next example illustrates the case where for the \( \theta^l \) types both IC and budget constraints are binding.

**Example 2** Consider the set of five treatments in table 2. Assume \( v_{0k} = \psi_k = c = 0 \) for each \( k \in K \). Hence treatment 1 has the highest CE score \( v/\delta \) and treatment 5 the lowest. Assume \( \beta^l = 4, \beta^h = 6.5 \). We start by assuming \( B = 0 \).

Contracts offered in the market are given in table 3. Note that we find \( \lambda^{il} \leq \lambda^{ih} \): as \( \theta^h \) spends her income on treatments with the highest CE scores, her marginal health return to income is higher; this is true in each income class \( i \in \{l, h\} \). The high income \( \theta^h \) type has lower values for \( \lambda \) and \( \nu \) than the low income \( \theta^l \) type as an increase in \( \beta \) relaxes both the budget and the IC constraints.

If the government increases its budget \( B > 0 \), it should subsidize treatment 2, not treatment 1 with the highest CE score. The reason is that treatment 2 is better targeted at type \( \theta^h \) because \( \rho^{hl}_{2} = 0 \) for \( i = l, h \). Hence each unit of the budget is spent on a type with \( \lambda = 4 \).\(^{26}\)

\(^{25}\)Although we do not use this result below, for completeness lemma 3 in the appendix derives necessary and sufficient conditions for \( \lambda^{il}, \nu^{il} \) to fall as income increases.

\(^{26}\)This is the value of \( \lambda^{lh} \) for a small increase in \( \beta^l \).
8. Optimal co-payments

Up till now we have taken $c^{ij}$ as given. In an RS style model, $\theta^l$ types face higher co-payments than $\theta^h$: in this way insurers separate high from low risk types. In an access to care model, there are two ways to separate types: higher co-payments and covering a different set of treatments. The following result derives sufficient conditions for $\theta^l$ to prefer higher co-payments than $\theta^h$ as in RS.

We use the following notation. We compare two arbitrary co-payments $c$ and $\bar{c}>c$. Let $\bar{V}^{ij}$ denote the (max.) utility for $ij$ from a contract with $c$. And $V^{ij}$ from a contract with co-payment $\bar{c}$. The sets in equation (3) are denoted by $K(\bar{K})$ for the low (high) co-payment.

**Proposition 3** Assume that (i) $K_i=\emptyset$ and (ii) for each $k \in K_n \cap K_e$ it is the case that $\theta^h_k = \theta^l_k$. Then we have the following

$$\bar{V}^{hh} - \bar{V}^{lh} \geq \bar{V}^{ih} - \bar{V}^{ih}$$

$$\bar{V}^{il} - \bar{V}^{il} \geq \bar{V}^{ih} - \bar{V}^{ih}$$

for each income group $i \in \{l, h\}$.

The proposition compares the change in utility under the high and low co-payment for different types. It shows high risk types with high income have a stronger preference for high co-payments than high risk types with low income. Further, within an income class, low risk types prefer the higher co-payment more than high risk types.

Increasing $c$ leads to the following trade-off at the individual level. As $c$ increases, more money $C^{ij} = c$ needs to be set aside to cover the co-payment $c$ and hence less money is available to spend on insurance. The opportunity cost of $C^{ij}$ is given by $\lambda^{ij}$. On the other hand, an increase in $c$ implies that more efficient treatments become available: $\bar{K}_c \cup \bar{K}_n \supseteq \bar{K}_e \cup \bar{K}_n$.
As implied by lemma 1, for a given c the $\lambda^j$ effect works in the direction of $\theta^j$ preferring higher c than $\theta^h$: opportunity cost of increasing c is lower. Similarly, a high income $\theta^h$ prefers higher c than a low income $\theta^h$ type.

This intuition is correct under two assumptions. Example 3 below shows that the result does not hold if assumption (i) is not satisfied. It is straightforward to see that a violation of assumption (ii) can lead to a situation where $\theta^h$ prefers a higher c than $\theta^l$. Assume that by increasing c only one thing happens: treatment k moves from $K_k$ to $K_n$ as agents have higher $C_{ij}$. If for this treatment $\theta^h_k > \theta^l_k = 0$, the increase in c has some value for $\theta^h$ (as k can now be financed directly out of $C^{(ij)}$). But because $\theta^l_k = 0$, this increase in c has not much value for the $\theta^l$ type. Hence, it can happen that $\theta^h$ has a higher preference for this increase in c than $\theta^l$.

The $\theta^h$ agent has an additional reason for high c: it facilitates separation from $\theta^h$ (standard RS argument). As can be seen in equation (IC), higher $c^l$ relaxes the IC constraint – for given sets $K_i, K_e, K_n$.

Example 3 Consider the treatments in table 4. We consider only one income class with $\beta = 2$. For treatment 1 we assume $\psi_1 = 0.95$. Covering treatment 1 destroys value unless the co-payment is at least equal to 1.9. Insurance covering 1 with c = 1.9 costs $\theta^h$: $0.2 \times (1 - 0.95)(10 - 1.9) + 1.9 = 2$ (approx.) and yields higher utility for $\theta^h$ than buying insurance covering treatments 2 and 3.

At c = 0 it is the case that $K_i = \{1\}$; hence assumption (i) in proposition 3 is not satisfied. Type $\theta^h$ strictly prefers $c = 1.9$ above $\underline{c}$. In this example, $\theta^h$ has a preference for higher c than $\theta^l$. In contrast to models in the line of RS, in an access to care model $\theta^h$ can have a higher co-payment (for all treatments) than $\theta^l$.

Finally, if co-payments differ between contracts, different types set aside different amounts of money $C^{ij}$. The following example illustrates that the result in lemma 1 does not necessarily hold if $C^{ij}$ differ.

Example 4 Consider the treatments in table 5 and assume $B = 0$. We focus on one income class with budget $\beta = 2$. Treatments 1 and 3 do not feature moral hazard problems, but treatment 2 does. We assume that $\psi_2 = 4/5$. Hence $\psi_2 \times 1 + (1 - \psi_2) \times 40 < \delta_2$: treatment 2 can only be covered with a co-payment equal to 1.

It is optimal for $\theta^h$ to buy insurance coverage for treatment 1. Hence, she chooses a contract with $c^h = 0, \rho^h_1 = 1$. She spends her whole budget on this. If she would get a small increase in income, she cannot afford the co-payment to make treatment 2 efficient and hence buys insurance for treatment 3: $\lambda^h = 3$.

Type $\theta^l$ buys coverage for treatment 2 with $c^l = 1, \rho^l_2 = 4/30$. Hence increasing $\theta^l$’s income marginally, allows her to extend her coverage of treatment 2: $\lambda^l = 4 > \lambda^h$.

Table 4: Value, costs and probabilities for treatments $K$ in example 3.

<table>
<thead>
<tr>
<th>K</th>
<th>$v_{1k}$</th>
<th>$v_{0k}$</th>
<th>$\delta_k$</th>
<th>$\theta_k^h$</th>
<th>$\theta_k^l$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>150</td>
<td>1.9</td>
<td>10</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>0</td>
<td>10</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>0</td>
<td>10</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>
If $c^{ij}$ are allowed to differ between types, it can happen that $\theta^l$ prefers a high co-payment and in this way makes a treatment $k$ available in an efficient way. If $\theta^h$ cannot afford this high co-payment, the health return to income can be higher for $\theta^l$ than for $\theta^h$. If $B > 0$, basic health insurance should target treatments that are mainly used by $\theta^l$ in this case because $\lambda^l > \lambda^h$.

Hence with the model above, redistribution from low income/high risk types to high income/low risk types is possible. It is an empirical question to determine which type has the highest return to an additional dollar spent on health care.

### 9. Policy implications

A number of countries feature a combination of public basic insurance and private supplementary insurance. A standard model based on moral hazard and adverse selection suggests that basic insurance should cover treatments that suffer most from adverse selection (Boone, 2014). However, many governments use cost effectiveness as criterion for inclusion in basic health insurance (Thomson et al., 2012). This suggests that the standard framework misses important aspects of the problem.

We introduced an access to care model. Indeed, in this model CE analysis plays a role. But basic insurance should not simply cover treatments with the highest CE scores. The main insight is that prevalence is important, though the planner’s objective function does not feature redistribution as an explicit goal.

When deciding on (changes in) the basic insurance package, two empirical issues need to be resolved. First, which types feature at the margin the highest health return to an increase in their budget. Second, which diseases (and their treatments) are prevalent among these types. Basic insurance should focus on coverage of these treatments.

For the first step, the data set that is used to estimate the risk adjustment model for a country is useful. It allows for the identification of types as in the risk adjustment system. The more detailed this system is, the better the planner can target. To illustrate, the Dutch risk equalization scheme features among others 38 age/gender groups, 20 groups based on past use of pharmaceuticals, 13 groups based on past diagnoses, 12 groups of social economic status (van de Ven and Schut, 2011). This creates a detailed grid of types for the Dutch population.

The question is the following. Consider a treatment $k$ that type $i$ currently covers with supplementary insurance. If treatment $k$ would be covered by basic insurance, this frees up financial resources for type $i$; how would these be spent? Would this be spent to create access for $i$ to a previously inaccessible treatment $\tilde{k}$; if so what is $\tilde{k}$’s return in terms of health? CE analysis is needed to determine $\tilde{k}$’s return. If, on the other hand, the freed up resources are spent on skiing holidays and/or cigarettes, the health return may be low. In this way, the

<table>
<thead>
<tr>
<th>$K$</th>
<th>$v_{1k}$</th>
<th>$v_{ok}$</th>
<th>$\delta_k$</th>
<th>$\theta^l_k$</th>
<th>$\theta^h_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>0</td>
<td>10</td>
<td>$\frac{3}{10}$</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>1</td>
<td>10</td>
<td>$\frac{3}{10}$</td>
<td>$\frac{3}{10}$</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>0</td>
<td>10</td>
<td>$\frac{3}{10}$</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5: Value, costs and probabilities for treatments $K$ in example 4.
government can determine the types with the highest health return per dollar spent on basic insurance.

For the second step, epidemiological studies are needed that identify the prevalence of diseases. Such studies do not only consider which conditions/symptoms tend to be correlated but also consider the prevalence of diseases among social classes, education groups, regions etc, (see e.g. Conen et al., 2009; Dalstra et al., 2005; Leyland, 2005; Van Loon et al., 1998). Hence this type of study can be done, although we are not aware of studies using the detailed type space used in risk adjustment.

Once the types with the highest health return are identified and the insured conditions that are more prevalent among these than among other types, the treatments of these conditions should be covered by basic health insurance. This gives the government the biggest health gain per dollar spent on basic health insurance.

Finally, should the government intervene in the health insurance market at all by offering basic insurance? In a standard insurance model, the welfare gain of such intervention is the reduction of the adverse selection inefficiency: under-insurance for low risk types. As argued by, for instance, Pauly (2008, pp. 121) the adverse selection inefficiency is rather small and hence competitive health insurance markets do a good job. Then there is not much need for government intervention. Indeed, if insurers are allowed to risk rate, the market outcome is efficient in the standard model and there is no reason for a welfare maximizing government to intervene.

However, in an access to care model the return to government intervention is much higher: it is related to the health benefit of treatments that low income/high risk agents cannot afford. This is true both with and without risk rating by insurers. Section 3 cited papers documenting that the uninsured in the US go without essential (and highly cost effective) health care. Then the return to government intervention is magnitudes higher than in a model without budget constraints.

One extension of the model is left for future research: the assumption that the agent’s condition is verifiable by a physician. There is an asymmetry here that we do not deal with in this paper. On the one hand, an agent can claim to have a disease and the physician can verify whether this is indeed the case. On the other hand, if an agent has two conditions she may decide to hide the symptoms of one of the conditions.27 Above we assume that the physician can verify all conditions that the agent suffers from. One can argue that it is easier for a physician to verify whether the patient suffers from a disease that she claims she has than to find the condition that the patient is hiding. Taking this asymmetry into account when designing the optimal insurance system is left for future research.

References


Ake Blomqvist. Chapter 12 - public-sector health care financing. In S. Glied and P. Smith, 27To illustrate, assume that treatment 1 for condition 1 has an affordable co-payment only if the agent does not suffer from condition 2 as well. An agent suffering from conditions 1 and 2, may then decide to hide condition 2 to get access to treatment 1.


A. Proof of results

Lemma 2 With second degree price discrimination, \( \theta^h \) and \( \theta^l \) are not pooled into one contract (in equilibrium) where at least one treatment \( k \) is covered with \( \theta_k^h > \theta_k^l \).

Proof of lemma 2 Suppose –by contradiction– that \( \theta^h \) and \( \theta^l \) with the same income \( \beta \) are pooled into one contract that covers \( (\rho_k > 0) \) at least one treatment \( k \) with \( \theta_k^h > \theta_k^l \). Then utility of type \( \theta^j \) can be written as

\[
 u^j = \beta + \sum_{k \in K^j} (1 - \psi_k) \rho_k (\theta^j_k (v_{1k} - C^{ij}) - \theta_k (\gamma_k - C^{ij})) \\
 + \sum_{k \in K^j} \rho_k (\theta^j_k (\psi_k v_{1k} + (1 - \psi_k) v_{0k} - C^{ij}) - \theta_k (\gamma_k - C^{ij})) \\
 + \sum_{k \in K^j} (1 - \psi_k) \theta^j_k (v_{1k} - \gamma_k)
\]

(A.1)

where \( \theta_k = \phi \theta^j_k + (1 - \phi) \theta_k^h \) is the pooled probability that someone (in this pool) needs treatment \( k \).

We claim that there exists a vector \( \hat{\rho} \) with \( \hat{\rho}_k \leq \rho_k \) for each \( k \) such that \( u^h = \hat{u}^h \) where

\[
 \hat{u}^h = \beta + \sum_{k \in K^j} (1 - \psi_k) \hat{\rho}_k (\theta^h_k (v_{1k} - C^{ij}) - \theta_k (\gamma_k - C^{ij})) \\
 + \sum_{k \in K^j} \hat{\rho}_k (\theta^h_k (\psi_k v_{1k} + (1 - \psi_k) v_{0k} - C^{ij}) - \theta_k (\gamma_k - C^{ij})) \\
 + \sum_{k \in K^j} (1 - \psi_k) \theta^h_k (v_{1k} - \gamma_k)
\]

(A.2)

This follows from the mean value theorem. At \( \hat{\rho}_k = \rho_k \) for each \( k \) we have that \( \hat{u}^h \geq u^h \). Further, \( \hat{\rho}_k = 0 \) for each \( k \) leads to \( \hat{u}^h \leq u^h \). By continuity there exists \( \hat{\rho} \) such that \( \hat{u}^h = u^h \). Further, we can choose \( \hat{\rho}_k < \rho_k \) for at least one treatment \( k \) where \( \theta_k^h > \theta_k^l \) due to the assumption that there is at least one treatment with \( \rho_k > 0 \) and \( \theta_k^h > \theta_k^l \).

In words, there exists a deviating contract for \( \theta^j \) such that \( \theta^h \) is indifferent between the original pooling contract and this deviation contract. That is, if this deviating contract is offered and priced at the \( \theta^j \) probabilities, type \( \theta^h \) will not switch to it. The following shows that \( \hat{u}^j > u^j \): \( \theta^j \) strictly prefers the deviation contract.

We can write

\[
 \hat{u}^j - u^j = \sum_{k \in K^j} (1 - \psi_k) [(\hat{\rho}_k - \rho_k) (\theta^j_k (v_{1k} - C^{ij}) - (\hat{\rho}_k \theta^j_k - \rho_k \theta_k) (\gamma_k - C^{ij}))] \\
 + \sum_{k \in K^j} [(\hat{\rho}_k - \rho_k) (\theta^j_k (v_{1k} - C^{ij}) - (\hat{\rho}_k \theta^j_k - \rho_k \theta_k) (\gamma_k - C^{ij}))]
\]

(A.3)

It is possible that \( \theta^h \) and \( \theta^l \) are pooled on a contract where the treatments covered feature \( \theta_k^h = \theta_k^l \); see example 1. Given their budget constraint, there is no information asymmetry for these types.

28
where treatments \( k \in K_{ij}^h \) drop out. It follows that

\[
(\hat{u}^h - u^h) - (\hat{u}^l - u^l) = \sum_{k \in K_{ij}^h} (1 - \psi_k)(\hat{\rho}_k - \rho_k)(\theta^h_k - \theta^l_k)(v_{1k} - C_{ij}) + \sum_{k \in K_{ij}^l} (\hat{\rho}_k - \rho_k)(\theta^h_k - \theta^l_k)(v_k - C_{ij})
\]

(A.4)

Because \( \hat{\rho}_k \leq \rho_k \) and \( \theta^h_k \geq \theta^l_k \) with at least one treatment where it is the case that \( \hat{\rho}_k < \rho_k \) and \( \theta^h_k > \theta^l_k \), we have that

\[
\hat{u}^l - u^l > \hat{u}^h - u^h = 0
\]

(A.5)

In words, \( \theta^l \) strictly prefers contract \( \hat{\rho} \) to contract \( \rho \). This implies that there exists a deviating contract that is strictly preferred by \( \theta^l \) while \( \theta^h \) prefers the pooling contract. Because \( \hat{u}^l > u^l \) the deviating contract can be offered in a profitable way. In equilibrium, types \( \theta^h \) and \( \theta^l \) cannot be pooled on a contract with \( \rho_k > 0 \) for at least one treatment with \( \theta^h_k > \theta^l_k \).

Next consider the IC constraints. Table A1 shows all potentially relevant constraints. We do not need to consider the diagonal elements where a type mimics itself. We argue that of the 12 remaining constraints in the table, at max. two are binding.

First, note that a high income type who mimics a low income type with the same \( \theta \) is actually not a problem in the sense that it does not affect the profitability of the contract (contract is based on the correct probabilities \( \theta \)).

Second, due to (5), each type spends her whole budget on insurance premium and copayment \( C_{ij} \). Hence, contracts can be chosen such that low income types cannot mimic high income types: they cannot afford the high income contracts. These are the “too expensive” entries in the table.

We just showed that \( \theta^l \) and \( \theta^h \) cannot be pooled into one contract. The same proof shows that \( \theta^l \) does not want to mimic \( \theta^h \). In particular, for each contract for \( \theta^h \) it is straightforward to find a contract for \( \theta^l \) which \( \theta^l \) strictly prefers while \( \theta^h \) does not.

Finally, type \( hh \) does not want to mimic \( ll \); this can be seen as follows:

\[
u^{hh} \geq \hat{u}^{lh} \geq u^{lh} \geq \hat{u}^{ll}
\]

where the first inequality follows from the fact that \( \theta^h \) with high income does not want to mimic \( \theta^h \) with low income (as there are treatments to insure which the low income type cannot afford); the second inequality follows from the fact that the high income type has higher income (even if \( hh \) mimics \( lh \) which gives \( hh \) utility \( \hat{u}^{lh} \), higher income leads to higher consumption and hence higher utility than \( u^{lh} \); the last inequality follows from \( IC_{lh,ll} \) and the fact that \( hh \) and \( lh \) have the same \( \theta^h \) and hence the same preferences.

Hence we only need to take into account two IC constraints: \( IC_{lh,ll} \) and \( IC_{hh,hl} \): the \( \theta^h \) type wants to mimic the \( \theta^l \) type with the same income.

\[Q.E.D.\]
**Proof of proposition 1** The government sets the values for $\gamma_k$. Given these values, the market offers equilibrium contracts that maximize $V^{ih}, V^{il}$ as given by equations (4) and (6) resp. However, it is easier to think of the planner as setting $\gamma_k, c^{ij}$ and $\rho_k$ to maximize welfare (7) while taking the same budget constraints and IC constraints into account that the market does. This allows us to use the envelop theorem as explained in appendix B.

Let $\mu$ denote the lagrange multiplier on constraint (GBC). Using the indicator functions defined in the main text, the first order condition for $\gamma_k$ can be written as

$$
\sum_i F^{i} \left[ \sum_j \phi_j^{ij} \left( I^{ij}_k \rho_k \lambda^{ij} + (1 - \psi_k) I_{\gamma_k \leq c^{ij}} \right) \right]
- \nu^{il} \phi_l^{ij} \left( \theta_l^{ij} I^{il}_k + (1 - \psi_k) I_{\gamma_k \leq c^{il}} \theta_k - \theta_k^{ih} (\rho_k^{ih} I_k^{ih} \lambda^{ih} + I_{\gamma_k \leq c^{ih}} (1 - \psi_k)) \right) = \mu \sum_i \sum_j F^{i} \phi_j^{ij} \left( I^{ij}_k \rho_k^{ij} + I_{\gamma_k \leq c^{ij}} (1 - \psi_k) \right)
$$

(A.6)

Note that $\gamma_k$ does not appear in this condition. In other words, the planner’s objective function is linear in $\gamma_k$. Thus we find a bang-bang solution for $\gamma_k$: it either equals $v_{0k}, c^{il}, c^{ih}, c^{il}, c^{hh}$ or $k$ is not covered: $\gamma_k = \delta_k$ (if the right hand side –costs– exceeds the left hand side –benefits). For the last treatment $k$ to be covered (before the government hits its budget constraint) the left hand side equals the right hand side. The government then sets $\gamma_k$ equal to the relevant $c^{ij}$ or $v_{0h}$ and the probability that the treatment is covered $\rho_k^g \in (0, 1]$ such that the budget $B$ is exactly spent. Rewriting equation (A.6), we see that the government ranks treatments using expression (10). It then covers the treatments that rank highest on this measure until the budget $B$ is spent. Q.E.D.

**Proof of proposition 2** Note that treatments with $\gamma_k \leq c^{ij}$ do not need to be insured as they are financed out of $C^{ij}$ when needed.

First, consider a type $\theta^h$: $\rho^{ih}$ is chosen to maximize (4). It is routine to see that the first order condition for $\rho^{ih}$ can be written as

$$
\rho^{ih} \begin{cases} 
= 1 & \text{if } s_k(\gamma_k, \rho^{ih}) > \lambda^{ih} \\
\in [0, 1] & \text{if } s_k(\gamma_k, \rho^{ih}) = \lambda^{ih} \\
= 0 & \text{if } s_k(\gamma_k, \rho^{ih}) < \lambda^{ih} 
\end{cases}
$$

(A.7)

Put differently, treatments are ranked according to $s_k(\gamma_k, c^{ij})$. Treatments with highest rank are insured ($\rho^{ih}_k = 1$) until the agent’s budget $\beta^i$ runs out.

For $\theta^l$, $\rho^{il}$ maximizes (6). With a similar reasoning as above, the first order condition then implies that treatments are ranked on the basis of

$$
s_k(\gamma_k, c^{il}) - \nu^{il} \left( \frac{\theta^h_k}{\theta^l_k} (s_k(\gamma_k, c^{il}) + 1) - 1 \right)
$$

(A.8)

It is routine to verify that this equation can be written as (15). Q.E.D.

**Proof of lemma 1** The inequality $\lambda^{ih} \geq \lambda^{ih}$ follows from the following observations. Both $\theta^h$ types rank treatments on the basis of $s_k(\gamma_k, c)$ and cover the highest ranking treatment by supplementary insurance. If $k$ denotes the last treatment covered by $ih$, then

$$
\lambda^{ih} = s_k(\gamma_k, c)
$$

(A.9)
As they have the same value for $C^{ij} = c^{ij} = c$ by assumption, they have (efficient) access to the same treatments. The $\theta^h$ type with the higher income can cover more treatments and hence has a lower $\lambda$.

Similarly, for $\theta^l$ type with $k$ the last treatment for which she buys insurance:

$$\lambda^u = s_k(\gamma_k, c) \left( 1 - \nu^l \left( \left( \frac{\theta^l_k}{\theta^l_h} - 1 \right) \frac{s_k(\gamma_k, c) + 1}{s_k(\gamma_k, c)} + 1 \right) \right)$$

(A.10)

Within the same income class $\beta^i$, we have $\lambda^{ih} \geq \lambda^{il}$. This follows from equations (A.9) and (A.10) in two steps. First, for given treatment $k$ the right hand side of (A.9) is at least as big as the right hand side of (A.10). Hence, if both types covered the same treatment at the margin, the claim is true. Second, $\theta^l$ covers a treatment at the margin which is lower in priority than $\theta^h$ because: (i) $\theta^l$ buys coverage for treatments $k$ at a premium based on probabilities $\theta^l_k \leq \theta^h_k$ and hence can cover more treatments, (ii) $\theta^l$ buys reduced (distorted) coverage of high priority treatments (based on (A.9)) to get separation from the $\theta^h$ contract and hence spends more money (than $\theta^h$) on treatments with lower priority $s_k$.

Q.E.D.

Lemma 3 Consider type $\theta^l$ with income $\beta^l$. Denote the two marginal treatments (last treatments to be covered) $k_1, k_2$ with shadow prices $\lambda^u, \nu^l$. For $\theta^l$ with income $\beta^h$ we denote the two marginal treatments $k_3, k_4$ where we order $k_3, k_4$ such that $v_{k_3}/\gamma_{k_3} \geq v_{k_4}/\gamma_{k_4}$. Then a necessary and sufficient condition for $\lambda^{hl} \leq \lambda^u$ is

$$\frac{\theta^l_{k_4}(v_{k_4} - c - \gamma_{k_4})}{\theta^h_{k_4}(v_{k_4} - c) - \theta^l_{k_4}(v_{k_4} - c - \gamma_{k_4})} - \frac{\theta^l_{k_3}(v_{k_3} - c - \gamma_{k_3})}{\theta^h_{k_3}(v_{k_3} - c) - \theta^l_{k_3}(v_{k_3} - c - \gamma_{k_3})} \leq \left( \frac{1}{\phi^l_{k_4} v_{k_4} - c - \gamma_{k_4}} - 1 \right) - \left( \frac{1}{\phi^h_{k_3} v_{k_3} - c - \gamma_{k_3}} - 1 \right) \lambda^u$$

(A.11)

A necessary and sufficient condition for $\nu^{hl} \leq \nu^l$ is

$$\frac{v_{k_3} - c}{\gamma_{k_3} - c - \theta^h_{k_3}} - \frac{v_{k_4} - c}{\gamma_{k_4} - c - \theta^h_{k_4}} \nu^l \geq \frac{v_{k_3} - \gamma_{k_3}}{\gamma_{k_3} - c} - \frac{v_{k_4} - \gamma_{k_4}}{\gamma_{k_4} - c}$$

(A.12)

Proof of lemma 3 Comparing two $\theta^l$ types (one with high, the other with low budget $\beta^l$) is a bit more involved than comparing two $\theta^h$ types because of the interaction of the budget and IC constraints. To see what happens with shadows prices $\nu, \lambda$ as budget $\beta^l$ increases, we derive the dual problem of an agent’s optimization problem. By assumption $C^{ij} = c$ for all types and hence the set $K_\gamma$ is the same for all types. It follows from proposition 1 that either $\gamma = c$ or $\gamma_k = \delta_k$. For treatments with $\delta_k > c$, assumption (1) implies that $\rho^{ij}_k = 0$ for each $k \in K_\gamma$; where $C^{ij} = c$ implies that the sets in (3) can be written without superscript $ij$.

Equation (6) can now be written as the a linear programming problem:

$$\max_{\rho_k \geq 0} \sum_{k \in K_\gamma} \rho_k (1 - \psi_k)(v_{k_1} - c)$$

(A.13)
subject to

\[ \rho_k \leq 1 \text{ for each } k \in K_e \]  
\[ \sum_{k \in K_e} \rho_k \theta_k^b (1 - \psi_k) (\gamma_k - c) \leq \beta^i \] (BC)
\[ \sum_{k \in K_e} \rho_k (1 - \psi_k) (\theta_k^b (v_{1k} - c) - \theta_k^i (\gamma_k - c)) \leq V^{ih} - \beta^i \] (IC)

where we made the following simplifications to ease notation. The right hand side of equation (BC) should be \( \beta^i - c \). But because \( C^{ij} = c \) for all types, we dropped the \( c \). Similarly, the objective function equals \( c \) plus the expression in (A.13), but we drop \( c \). As all types have the same set \( K_n \), we dropped the term \( \sum_{k \in K_n} (1 - \psi_k) \theta_k^b (v_{1k} - \gamma_k) \) in equation (IC).

The dual of problem (A.13) is given by (see, for instance, Brinkhuis and Tikhomirov, 2005)

\[
\min_{\zeta_k, \lambda, \nu \geq 0} \sum_{k \in K_e} \zeta_k + \beta^i \lambda + (V^{ih} - \beta^i) \nu
\]

subject to

\[
\zeta_k + \theta_k^i (1 - \psi_k) (\gamma_k - c) \lambda + (1 - \psi_k) (\theta_k^b (v_{1k} - c) - \theta_k^i (\gamma_k - c)) \nu \geq \theta_k^i (1 - \psi_k) (v_{1k} - c) \tag{A.16}
\]

where \( \zeta_k \) is the shadow price corresponding to the constraint \( \rho_k \leq 1 \), \( \lambda \) corresponds to (BC) and \( \nu \) to (IC).

For \( \beta^i \geq 0 \) close to 0, \( V^{ih} - \beta^i \geq 0 \) is close to zero as well. It is then optimal to choose \( \zeta_k = 0 \) for each \( k \) in the optimization problem (A.15). Hence, \( \lambda \) and \( \nu \) are chosen such that all inequalities (A.16) hold with \( \zeta_k = 0 \). If \( \lambda, \nu > 0 \), they are determined by 2 treatments \( k_1, k_2 \) such that \( \zeta_{k_1} = \zeta_{k_2} = 0 \) and equation (A.16) holds with equality for \( k_1 \) and \( k_2 \).

As \( \beta^i \) and \( V^{ih} - \beta^i \) increase, at some point it becomes optimal to lower \( \lambda, \nu \) such that \( \zeta_{k_1}, \zeta_{k_2} > 0 \) and \( \lambda, \nu > 0 \) are determined by two treatments \( k_3, k_4 \) with \( \zeta_{k_3} = \zeta_{k_4} = 0 \) and (A.16) holding with equality for these two treatments. In the primal problem this corresponds to \( \beta^i \) high enough that \( \rho_{k_1} = \rho_{k_2} = 1 \).

This is illustrated in figure 2 based on example 2. The solid lines correspond to equations (A.16) holding with equality and \( \zeta_k = 0 \) for each of the 5 treatments \( k \in K \):

\[
\lambda = \frac{v_{1k} - c}{\gamma_k - c} - \left( \frac{\theta_k^b v_{1k} - c}{\theta_k^i \gamma_k - c} - 1 \right) \nu \tag{A.17}
\]

The dashed lines correspond to the slope of the objective function (A.15) in \( (\nu, \lambda) \) space. For low values of \( \beta^i \) and \( V^{ih} - \beta^i \), it is optimal to set \( \zeta_k = 0 \) for each \( k \). For \( (\nu^b, \lambda^b) \) all inequalities (A.16) are satisfied and they are binding for treatments \( k_1 = 1 \) and \( k_2 = 4 \). At higher values \( \beta^b \) and \( V^{ih} - \beta^b \), it is optimal to set \( \zeta_1, \zeta_4 > 0 \) and the other \( \zeta_k = 0 \). At \( (\nu^{bl}, \lambda^{bl}) \) the binding constraints (A.16) correspond to treatments \( k_3 = 3 \) and \( k_4 = 5 \).

\[29\]If either \( \lambda = 0 \) or \( \nu = 0 \), the other shadow price is determined by \( \zeta_{k_1} = 0 \) and (A.16) holding with equality for \( k_1 \). As we are interested in the case where both (BC) and (IC) are binding, we focus on \( \lambda, \nu > 0 \). It is straightforward to extend the analysis below for the case where one shadow price equals 0.
We can now derive a necessary and sufficient condition for $\lambda^{hl} \leq \lambda^{ll}$. Consider the treatment with the lower intercept $(v_k - \gamma_k)/(\gamma_k - c)$ in equation (A.17). The intersection of this line with $\lambda = \lambda^{ll}$ should lie to the left of the intersection of the $k_3$-line with $\lambda$:

$$\frac{\theta_{k_4}^l (v_{k_4} - \gamma_{k_4})}{\theta_{k_4}^h (v_{k_4} - c)} - \frac{1}{\theta_{k_4}^h (v_{k_4} - c) - \gamma_{k_4} - c} \leq \frac{\theta_{k_3}^l (v_{k_3} - \gamma_{k_3})}{\theta_{k_3}^h (v_{k_3} - c) - \gamma_{k_3} - c} - \frac{1}{\theta_{k_3}^h (v_{k_3} - c) - \gamma_{k_3} - c}$$

This can be written as equation (A.11).

Similarly, we can derive the necessary and sufficient condition for $\nu^{hl} \leq \nu^{ll}$. In words, at $\nu^{ll}$ the line corresponding to $k_3$ should lie below the line corresponding to $k_4$. Then the intersection point of the two lines is to the left of $\nu^{ll}$. This can be written as:

$$\frac{v_{k_3} - \gamma_{k_3}}{\gamma_{k_3} - c} - \left(\frac{\theta_{k_4}^h (v_{k_4} - c)}{\theta_{k_4}^h (v_{k_4} - c) - \gamma_{k_4} - c} - 1\right) \nu^{ll} \leq \frac{v_{k_3} - \gamma_{k_3}}{\gamma_{k_3} - c} - \left(\frac{\theta_{k_4}^h (v_{k_4} - c)}{\theta_{k_4}^h (v_{k_4} - c) - \gamma_{k_4} - c} - 1\right) \nu^{ll}$$

which is equivalent to equation (A.12).

Proof of proposition 3 In order to derive a “marginal” condition, we introduce the following “lottery contract”. Fix two values $c > \bar{c} \geq 0$. The agent reserves an amount $c_\alpha = \alpha \bar{c} + (1 - \alpha) c$ for co-payments. This $c_\alpha$ is the price of a lottery ticket which is linked to health insurance in the following way. Before learning the realization of the lottery, the agent buys insurance coverage $\rho_k$. After the insurance is bought –but before treatment is used– the payoff from the lottery is realized. The lottery pays out $\bar{c}(\bar{c})$ with probability $\alpha(1 - \alpha)$. If the lottery pays $\bar{c}(\bar{c})$, the agent faces co-payment $\bar{c}(\bar{c})$ when using treatment $k$. The relevant treatment
sets are denoted $\bar{K}_{e,i,n}(K_{e,i,n})$. For given $\alpha$, if there are no IC constraints, agent $ij$ solves

$$V^{ij}_\alpha = \max_{\rho_k} \beta^i + \alpha \left( \sum_{k \in K_e} \theta^i_k (1 - \psi_k) \rho_k (v_{1k} - \gamma_k) + \sum_{k \in K_i} \theta^i_k \rho_k (v_k - \gamma_k) + \sum_{k \in K_n} \theta^i_k (1 - \psi_k) (v_{1k} - \gamma_k) \right)$$

$$+ (1 - \alpha) \left( \sum_{k \in K_e} \theta^i_k (1 - \psi_k) \rho_k (v_{1k} - \gamma_k) + \sum_{k \in K_i} \theta^i_k \rho_k (v_k - \gamma_k) + \sum_{k \in K_n} \theta^i_k (1 - \psi_k) (v_{1k} - \gamma_k) \right)$$

$$- \lambda^{ij} \left( \alpha \left( \sum_{k \in K_e} \theta^i_k (1 - \psi_k) \rho_k (\gamma_k) + \sum_{k \in K_i} \theta^i_k \rho_k (\gamma_k) \right) \right)$$

$$+ (1 - \alpha) \left( \sum_{k \in K_e} \theta^i_k (1 - \psi_k) \rho_k (\gamma_k) + \sum_{k \in K_i} \theta^i_k \rho_k (\gamma_k) \right) - (\beta - c_\alpha) \right)$$

(A.20)

Clearly, $c = \bar{c}$ leads to higher utility than $c = \underline{c}$ if and only if $V^{ij}_{\alpha = 1} > V^{ij}_{\alpha = 0}$. But because the optimal $\rho_k$'s are different at $\alpha = 1$ than at $\alpha = 0$, the condition $V^{ij}_{\alpha = 1} > V^{ij}_{\alpha = 0}$ does not easily convey the relevant trade-offs. Therefore, we consider $\alpha \in [0, 1]$ and evaluate the marginal effect $da$. Then we can use (with a slight abuse of notation) $V^{ij}_{\alpha = 1} - V^{ij}_{\alpha = 0} = \int_0^1 dV^{ij}_\alpha / d\alpha$. The envelope theorem implies

$$\frac{dV^{ij}_\alpha}{d\alpha} = \sum_{k \in K_{e,i,n}} \theta^i_k \rho^j_k (\psi_k (\gamma_k - v_{0k}) + \lambda^{ij} (\gamma_k - \underline{c} - (1 - \psi_k) (\gamma_k - \bar{c})))$$

$$+ \sum_{k \in K_{e,i,n}} \theta^i_k (1 - \psi_k) ((1 - \rho^j_k) (v_{1k} - \gamma_k) + \rho^j_k \lambda^{ij} (\gamma_k - \underline{c}))$$

$$+ \sum_{k \in K_{e,i,n}} \theta^i_k ((1 - \psi_k) (1 - \rho^j_k) (v_{1k} - \gamma_k) + \rho^j_k (\gamma_k - v_{0k}) + \rho^j_k \lambda^{ij} (\gamma_k - \underline{c}))$$

$$- \lambda^{ij} (\bar{c} - \underline{c})$$

$$= \sum_{k \in K_{e,i,n}} \theta^i_k (1 - \psi_k) ((1 - \rho^j_k) (v_{1k} - \gamma_k - \lambda^{ij} (\gamma_k - \underline{c})) - \lambda^{ij} (\bar{c} - \gamma_k)))$$

$$- (1 - \sum_{k \in K_{e,i,n}} \theta^i_k (1 - \psi_k)) \lambda^{ij} (\bar{c} - \underline{c})$$

(A.21)

The first equality follows from the following two observations. First, $\bar{K}_j \supseteq K_{e,i,n}$ for $j = e, n$ because $\bar{c} > \underline{c}$. Second, treatments $k \in K_{e,i,n} \cap K_j$ drop out in the expression for $dV^{ij}_\alpha / d\alpha$. For the second equality, we use $K_{e,i,n} = \emptyset$.

The remaining summation is over treatments $k$ that are used efficiently at $\underline{c}$ and can be financed without insurance at $\bar{c}$. If such a treatment is not insured at $\underline{c}$ (i.e. $\rho^j_k = 0$), then the increase in $c$ makes the treatment available which leads to utility gain $v_{1k} - \gamma_k$ when used. If the treatment was already covered by insurance (i.e. $\rho^j_k = 1$), it is used both at $\underline{c}$ and at $\bar{c}$. The only effect of increasing $\alpha$ is then to save money on the insurance premium $\theta^j_k (1 - \psi_k) (\gamma_k - \underline{c})$. 


This positive effect of $d\alpha > 0$ needs to be weighed against the increase in $c_\alpha$ as $\alpha$ increases. The cost of this increase $dc_\alpha/d\alpha = \bar{c} - \underline{c}$ is valued at the marginal utility of income $\lambda^{ij}$. Higher $\lambda^{ij}$ makes an increase in $\alpha$ less attractive. However, in order to compare $dV^{\alpha ij}/d\alpha$ across types, we need to take the different values of $\rho_k^{ih}$ into account as well.

First, we compare $dV^{\alpha h}_\alpha/d\alpha$ with $dV^{\alpha h}_\lambda/d\alpha$. Using (A.21), we write

$$\frac{d(V^{\alpha h}_\alpha - V^{\alpha h}_\lambda)}{d\alpha} = \left(1 - \sum_{k \in \tilde{K}_n \cap \tilde{K}_e} \theta_k(1 - \psi_k)\left(\lambda^{\alpha h} - \lambda^{\alpha h}(\bar{c} - \underline{c})\right) + \sum_{k \in \tilde{K}_n \cap \tilde{K}_e} \theta_k(1 - \psi_k)\left[\left((1 - \rho_k^{\alpha h})(\nu_{1k} - \gamma_k - \lambda^{\alpha h}(\gamma_k - \underline{c})\right) - \lambda^{\alpha h}(\bar{c} - \gamma_k)\right) - \left((1 - \rho_k^{\alpha h})(\nu_{1k} - \gamma_k - \lambda^{\alpha h}(\gamma_k - \underline{c})\right) - \lambda^{\alpha h}(\bar{c} - \gamma_k)\right]\right) \geq 0$$

where the inequality follows from the following considerations. First, it follows from lemma 1 that $\lambda^{\alpha h} - \lambda^{\alpha h} \geq 0$. As $k \in \tilde{K}_n$ implies $\bar{c} > \gamma_k$, a sufficient condition for the inequality to hold is

$$(1 - \rho_k^{\alpha h})(\nu_{1k} - \gamma_k - \lambda^{\alpha h}(\gamma_k - \underline{c})\right) \geq (1 - \rho_k^{\alpha h})(\nu_{1k} - \gamma_k - \lambda^{\alpha h}(\gamma_k - \underline{c})\right)$$

As both $\theta_k$ types use the same ordering (14) of treatments and the $\beta_k$ type has more money to spend, we find that $\rho_k^{\alpha h} \geq \rho_k^{\alpha h}$.

Next consider the comparison within income class $\beta_k$ of $\theta_k$ and $\theta_k^{\alpha}$. We want to show that $\theta_k^{\alpha}$ has a stronger preference for higher $c$ than $\theta_k$. We do this in two steps. First, we assume that no IC is binding. Then we consider the case where $\nu_{1k} > 0$. With $\nu_{1k} = 0$ we have

$$\frac{d(V^{\alpha l}_\alpha - V^{\alpha l}_\lambda)}{d\alpha} = \left(1 - \sum_{k \in \tilde{K}_n \cap \tilde{K}_e} \theta_k(1 - \psi_k)\left(\lambda^{\alpha l} - \lambda^{\alpha l}(\bar{c} - \underline{c})\right) + \sum_{k \in \tilde{K}_n \cap \tilde{K}_e} \theta_k(1 - \psi_k)\left[\left((1 - \rho_k^{\alpha l})(\nu_{1k} - \gamma_k - \lambda^{\alpha l}(\gamma_k - \underline{c})\right) - \lambda^{\alpha l}(\bar{c} - \gamma_k)\right) - \left((1 - \rho_k^{\alpha l})(\nu_{1k} - \gamma_k - \lambda^{\alpha l}(\gamma_k - \underline{c})\right) - \lambda^{\alpha l}(\bar{c} - \gamma_k)\right]\right) \geq 0$$

where we have used the assumption that $\rho_k^{\alpha l} = \theta_k^{\alpha l} = 0$ for each $k \in \tilde{K}_n \cap \tilde{K}_e$. Using lemma 1, we have $\lambda^{\alpha l} \geq \lambda^{\alpha l}$. As above, a sufficient condition for the inequality to hold is

$$(1 - \rho_k^{\alpha l})(\nu_{1k} - \gamma_k - \lambda^{\alpha l}(\gamma_k - \underline{c})\right) \geq (1 - \rho_k^{\alpha l})(\nu_{1k} - \gamma_k - \lambda^{\alpha l}(\gamma_k - \underline{c})\right)$$

Note that the right hand side of this inequality is non-positive. Hence, with $v_{1k} - \gamma_k - \lambda^{\alpha l}(\gamma_k - \underline{c}) \geq 0$ the inequality is satisfied. Consider $v_{1k} - \gamma_k - \lambda^{\alpha l}(\gamma_k - \underline{c}) < 0$ (which implies $\rho_k^{\alpha l} = 0$). Then $\lambda^{\alpha l} \geq \lambda^{\alpha l}$ implies that $\rho_k^{\alpha l} = 0$ and

$$v_{1k} - \gamma_k - \lambda^{\alpha l}(\gamma_k - \underline{c}) < v_{1k} - \gamma_k - \lambda^{\alpha l}(\gamma_k - \underline{c})$$

and again the inequality above is satisfied.
Now we turn to the case where \( \nu^d > 0 \). This introduces the following additional term:

\[
+ \nu^d \left( \beta^i + \alpha \left( \sum_{k \in K_n} \theta^d_k (1 - \psi_k) \rho_k (v_{1k} - \gamma_k) + \sum_{k \in K_i} \theta^d_k (1 - \psi_k) \rho_k (v_{1k} - \gamma_k) \right) + (1 - \alpha) \left( \sum_{k \in K_n} \theta^b_k (1 - \psi_k) \rho_k (v_{1k} - \gamma_k) + \sum_{k \in K_i} \theta^b_k (1 - \psi_k) \rho_k (v_{1k} - \gamma_k) \right) - \tilde{V}^{ih} \right)
\]

(A.22)

where the value for \( ih \) of mimicking \( il \) is given by

\[
\tilde{V}^{ih}_\alpha = \beta^i + \alpha \left( \sum_{k \in K_n} (1 - \psi_k) \rho^d_k (\theta^d_k v_{1k} - \theta^d_k \gamma_k - \Delta \theta_k \bar{c}) + \sum_{k \in K_i} (1 - \psi_k) \rho^d_k (\theta^d_k v_{1k} - \theta^d_k \gamma_k - \Delta \theta_k \bar{c}) \right) + (1 - \alpha) \left( \sum_{k \in K_n} (1 - \psi_k) \rho^b_k (\theta^b_k v_{1k} - \theta^b_k \gamma_k - \Delta \theta_k \bar{c}) + \sum_{k \in K_i} (1 - \psi_k) \rho^b_k (\theta^b_k v_{1k} - \theta^b_k \gamma_k - \Delta \theta_k \bar{c}) \right)
\]

(A.23)

Hence we find that

\[
\frac{d\tilde{V}^{ih}_\alpha}{d\alpha} = \sum_{k \in K_n \cap K_e} \theta_k (1 - \psi_k) (1 - \rho^d_k)(v_{1k} - \gamma_k)
\]

Hence, differentiating equation (A.22) with respect to \( \alpha \) gives

\[
\nu^d \left( \sum_{k \in K_n \cap K_e} \theta_k (1 - \psi_k) (\rho^d_k - \rho^b_k)(v_{1k} - \gamma_k) \right) \geq 0
\]

where the inequality follows from the following considerations. First, if \( \rho^d_k = 0 \), the inequality holds. Second, consider \( \rho^d_k = 1 \). This implies that treatment \( k \) ends up high in the ranking (13). But because \( \theta^d_k = \theta^d_k \) for the treatments in this summation, the ranking in (15) is high as well and \( \theta^d \) has more money to spent as insurance is cheaper for the better risk. Thus we see that \( \rho^b_k = 1 \) implies that \( \rho^d_k = 1 \) and the inequality is satisfied. Intuitively, as \( \alpha \) increases, treatment \( k \) is more likely to end up in \( K_n \) instead of \( K_i \); this makes mimicking less attractive. Q.E.D.

B. Using the envelop theorem

When we consider the effects of the government’s choice \( \gamma_k \) in section 5, we focus on the direct effects; ignoring the effects of \( \gamma_k \) on \( \rho^d_k, c^i, c^j \). This is due to the envelop theorem. However, it may not be obvious that the envelop theorem can be applied in this context. The
following proposition derives an envelop theorem result in the context of IC constraints. We first introduce some notation.

Let \( u_i(c_i, \gamma) \) denote type \( i \in \{l, h\} \)'s utility as a function of \( i \)'s choice vector \( c_i \) and parameter vector \( \gamma \) set by the government. The vector \( c_i \) now incorporates the variables \( \rho_{ik}, c_{ij}, C^j \). We have \( c_i \in C \) where the set \( C \) includes constraints like \( \rho_k \in [0, 1] \). The function \( u_i \) can be a Lagrangian with budget constraint and multiplier \( \lambda_i \). If type \( j \) mimics \( i \), \( j \)'s utility is written as \( \hat{u}_j(c^j, \gamma) \). Finally, \( \nu_i \) denotes the Lagrange multiplier on \( j \)'s IC constraint not to mimic \( i \).

In our analysis, the following assumption is satisfied.

**Assumption 1** Type \( h \) may want to mimic \( l \), but \( l \) does not mimic \( h \).

In our model this is correct within an income class. Types do not mimic across income class.

Then we have

\[
U_h(\gamma) = \max_{c^h \in C} u_h(c^h, \gamma) \tag{B.24}
\]

\[
U_l(\gamma) = \max_{c^l \in C} u_l(c^l, \gamma) - \nu_l \left( U_h(\gamma) - \hat{u}_h(c^l, \gamma) \right) \tag{B.25}
\]

Now we have the following envelop result.

**Proposition 4** With assumption 1 we find that

\[
\frac{dU_h(\gamma)}{d\gamma} = \frac{\partial u_h(c^h, \gamma)}{\partial \gamma} \tag{B.26}
\]

\[
\frac{dU_l(\gamma)}{d\gamma} = \frac{\partial u_l(c^l, \gamma)}{\partial \gamma} - \nu_l \left( \frac{dU_h(\gamma)}{d\gamma} - \frac{\partial \hat{u}_h(c^l, \gamma)}{\partial \gamma} \right) \tag{B.27}
\]

Note that there are two possibilities here for an element \( c^h_j \) in vector \( c^h \). First, \( c^h_j \) is an interior solution. Then (B.26) holds because \( \partial u_h/\partial c^h_j = 0 \). Second, \( c^h_j \) is a corner solution. Then generally speaking \( \partial u_h/\partial c^h_j \neq 0 \), but a small change in \( \gamma \) will not affect \( c^h_j \): \( dc^h_j/d\gamma = 0 \).