

The Effects of Capping Co-Insurance Payments

Westerhout, Ed; Folmer, Kees

Document version:

Early version, also known as pre-print

Publication date:

2018

[Link to publication](#)

Citation for published version (APA):

Westerhout, E., & Folmer, K. (2018). *The Effects of Capping Co-Insurance Payments*. (CentER Discussion Paper; Vol. 2018-050). CentER, Center for Economic Research.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.



No. 2018-050

THE EFFECTS OF CAPPING CO-INSURANCE PAYMENTS

By

Ed Westerhout, Kees Folmer

22 November 2018

ISSN 0924-7815
ISSN 2213-9532

The effects of capping co-insurance payments

Ed Westerhout and Kees Folmer

Ed Westerhout: CPB Netherlands Bureau for Economic Policy Analysis, P.O. Box 80510, 2508 GM Den Haag, THE NETHERLANDS, ewmtw@cpb.nl; Tilburg University, Tilburg School of Economics and Management, Warandelaan 2, 5037 AB Tilburg, E.W.M.T.Westerhout@uvt.nl.
Kees Folmer: CPB Netherlands Bureau for Economic Policy Analysis, P.O. Box 80510, 2508 GM Den Haag, THE NETHERLANDS, C.E.Folmer@cpb.nl.

Abstract:

It is common to limit the cost sharing in health insurance schemes by a cap on co-insurance payments. This paper derives the economic and welfare effects of such a cap, adopting a model of which two features are crucial. First, health care demand is price-elastic. Second, demand is less elastic the worse the consumer's health status. The paper derives that a cap induces optimizing health insurers to raise the co-insurance rate. This raises welfare in the aggregate, but part of the consumers do not share in this welfare gain. In particular, those with health spending close to the level at which co-insurance payments reach their maximum level suffer large welfare losses. We adopt a 3-state model to derive our results and a continuous-state model for a numerical illustration.

Keywords: Coinsurance, Health Insurance, Cap on Coinsurance Payments, Moral Hazard

JEL codes: D60, H21, I18

Acknowledgements

An earlier version of this paper has been presented at the 2010 IIPF Conference. Thanks are due to the participants at this meeting, in particular our discussant Jan-Maarten van Sonsbeek. Further, we thank Roel Beetsma, Eddy van Doorslaer, Bas Jacobs, Thomas McGuire, Erik Schut and Ward Romp for valuable comments on earlier versions of this paper. Any remaining errors are the responsibility of the authors.

1 Introduction

Throughout the world, health insurance schemes feature capped co-insurance.¹

This may sound obvious since without co-insurance moral hazard would increase the demand for health care too much. By the same token, the use of a cap is obvious as not using it might impose too much risk on consumers.

Indeed, capped insurance may be better in welfare terms than no insurance or full insurance. This accords with a considerable literature that suggests that optimal insurance is a form of partial insurance, striking a balance between the welfare loss from moral hazard and the welfare gain from risk sharing (Pauly 1968; Zeckhauser 1970; Besley 1988; Manning and Marquis 1996; Blomqvist 1997).²

Surprisingly, there is little literature on the effects of imposing a cap on co-insurance payments. This paper explores the economic, distributional and welfare effects of introducing a cap. It analyses the features of an optimal co-insurance scheme in a world in which insurers can apply a co-insurance rate up to a certain maximum and full insurance beyond that point. Two features of the demand for health care services will prove to be relevant: demand is price-

¹ For example, in the United States, the Affordable Care Act requires that all health plans limit participant out-of-pocket (OOP) maximums. The Netherlands use deductibles in their insurance scheme, which - by construction - maximize OOP payments to the amount of the deductible. Insurance schemes with caps on co-payments can also be found in other countries, such as Austria, Finland, Iceland, Ireland, Norway, Sweden and Switzerland (OECD 2015).

² There is a large literature on optimal coinsurance schemes. For example, Besley (1988) shows that optimal coinsurance rates will generally be different for different types of health care, Ellis and McGuire (1990) and Ma and McGuire (1997) point to the interaction between supplier cost sharing and coinsurance policies and Newhouse (2006) and Chandra, Gruber and McKnight (2010) underline the role of offset effects. Glazer and McGuire (2012) discusses the welfare impact of these offset effects and the role of insurance. Nyman (1999) stresses that without health insurance, some (very expensive) medical services would not be affordable. Newhouse (2006) explores the role of a lack of will power on part of patients and Pauly and Blavin (2008) that of information imperfections on part of patients.

elastic, and the (absolute value of the) price elasticity is decreasing with health status. Hence, the worse someone's health, the less price-elastic his or her demand for health care services will be.

We derive first that, generally, the optimal co-insurance rate can take any positive value and is not restricted to be equal or smaller than unity. The reason is that redistribution may imply both a lower or a higher co-insurance rate. If we impose the institutional requirement that the co-insurance rate cannot exceed one, our analysis thus implies that in some cases optimal policies will feature a deductible. Second, for the case in which the optimal co-insurance rate is strictly between zero and one, we restate the result established by many others that the optimal co-insurance rate is increasing in the price elasticity of health care demand. Thirdly, we will show that the optimal maximum of co-insurance payments is finite. Hence, any co-insurance scheme without an upper bound to co-insurance payments must be suboptimal. Fourthly, due to the properties of health care demand, the introduction of a cap on insurance increases the price-elasticity of health care spending below the cap, and thus enforces health insurers to raise the co-insurance rate. This is interesting in itself, but it also implies that the redistributive effects of the introduction of a cap on co-insurance payments are very uneven. In particular, consumers whose spending on health care is just below the imposed cap may face a steep increase in co-insurance payments due to the rising optimal co-insurance rate.

We use a simple model to derive our results. In this model there is just one medical product, and consumers are allocated to three possible health states. In the first costs are zero. In the second state the out-of pocket price is positive,

but costs are below a co-payment maximum. In the third state health expenditures exceed that maximum and the out-of-pocket price is zero. Subsequently, we present a second model that extends this simple framework to one that allows for a continuum of health states and two medical products. Numerical simulations with this extended model suggest that the results from the simple model are non-trivial.

There are several contributions to the literature that relate to our work. Like us, Keeler, Newhouse and Phelps (1977), Ellis (1986), Manning and Marquis (1996) and Kowalski (2012) use models with nonlinear budget constraints. However, they do not explore the implications for optimal cost sharing. Finkelstein and McKnight (2008) and Engelhardt and Gruber (2011) do assess the welfare implications of moral hazard and risk reduction, but they apply separate models to assess the changes in moral hazard and risk reduction, and the associated welfare effects. This approach may be too rough to find an optimal scheme: we prefer to assess these changes simultaneously. Closely related are also Besley (1988) and Goldman and Philipson (2007) who show theoretically the relation between the optimal co-insurance rate and the price elasticity of demand, and Einav, Finkelstein and Polyakova (2016) who present empirical evidence for this relationship in the case of prescription drugs. Also related is the work of Drèze and Schokkaert (2013) on the optimality of deductibles when demand for health care is subject to ex post moral hazard, and Blomqvist (1997) on optimal nonlinear health insurance schemes.

The structure of our paper is as follows. Section 2 elaborates the three-state model and section 3 explains the results graphically. Section 4 discusses the

continuous-state model. Section 5 presents the results of numerical simulations with the latter model. Section 6 concludes.

2 Optimal co-insurance in a three-state world

In this section we consider a model that distinguishes three states for health care demand: zero demand, positive and price-elastic demand and positive and price-inelastic demand. Basically, the analysis in this section confirms the result derived by many others (Pauly (1968), Zeckhauser (1970), Besley (1988) and Goldman and Philipson (2007)) that the optimal co-insurance rate is increasing in the price elasticity of health care demand (in absolute terms). Here, we add two things, namely that this result does not hold generally and that there is also an optimal co-insurance maximum that relates to the optimal co-insurance rate.

2.1 The model

We adopt a representative-agent approach. Hence, *ex ante* all consumers are alike. *Ex post* however, heterogeneity steps in as people who are identical from an *ex ante* view encounter different shocks to their health. In other respects, our model is as stylized as possible: we abstain from modelling explicitly the role of suppliers, information imperfections and any offset effects.

Ellis and McGuire (1990) and Eggleston (2000) use a quadratic specification to describe the relation between utility and medical care. We adopt their approach, primarily for two reasons. First, quadratic utility implies that the

marginal utility of health drops to zero for some finite amount of care. This allows us to describe demand also in case of a zero out-of-pocket price, which is quite common in health care. Second, there is empirical evidence that health care demand is less price-elastic for higher amounts of health care spending. For example, Wedig (1988) already has found that the price elasticity of health care demand is smaller the worse the health status of the patient. Newhouse and the Insurance Experiment Group (1993), Van Vliet (2001), Zhou *et al.* (2011) and Chandra, Gruber and McKnight (2014) present evidence that the demand for expensive inpatient hospital services is less price-elastic than that for less expensive alternative medical services. Moreover, Strombom, Buchmueller and Feldstein (2002) and Buchmueller (2006) have found that, compared to workers, the demand for health insurance by retirees is little price-elastic; the same holds true for the demand by people who are older and who have been recently hospitalized or diagnosed with cancer. The quadratic specification is consistent with this evidence: it implies that the price elasticity of health care demand is decreasing in health status and spending. Moreover, it also implies that the price elasticity of health care demand is increasing in the co-insurance rate (see Appendix A). This is backed by empirical evidence as well (Phelps and Newhouse 1974).

The quadratic form has a drawback as it implies that risk aversion is increasing in income. In our analysis, this argument has little weight since we do not account for income heterogeneity. In section 5 we will explore alternative co-insurance schemes that differ in terms of premium. We will see

that the implied differences in disposable income are too small to exert important effects upon risk aversion.

If u denotes the consumer's direct utility, z the consumption of health care, c consumption of non-medical services and y consumer's gross income, then we have the following utility function:

$$u = c - \frac{1}{2}\beta c^2 + \gamma z - \frac{1}{2}\delta z^2 \quad (1)$$

$$0 \leq \beta < 1/y, \quad \delta > 0$$

The range of the parameter β assures that the marginal utility of non-medical consumption, $1 - \beta y$, is always positive. The parameter γ differs across states of health. This reflects patient heterogeneity in terms of health status. Health status affects utility as bad health increases the efficacy of medical care. Indeed, (1) implies that the marginal utility of health care consumption $\gamma - \delta z$ is increasing in γ . Preferences for medical and non-medical consumption in (1) are separable. Finkelstein, Luttmer and Notowidigdo (2013) have shown empirical evidence for a non-separable form in which the marginal utility of non-medical consumption declines when health deteriorates. This has a similar effect as our specification, namely shifting the consumption basket away from non-medical goods and towards health care when health deteriorates.

The rate of co-insurance is denoted as b with $0 \leq b \leq 1$. We use t to denote the producer price of medical services, so bt measures the out-of-pocket price. On account of the maximum to co-insurance payments m , the budget constraint of the consumer is nonlinear:

$$\begin{aligned}
c &= y - p - btz & 0 \leq z \leq \frac{m}{bt} \\
c &= y - p - m & \frac{m}{bt} \leq z
\end{aligned}
\tag{2}$$

Here, p denotes the health insurance premium.

The consumer maximizes (1), subject to (2), given the value of the parameter γ that reflects his health status. Principally, this problem is non-standard due to the two-part structure for the price of medical services. We will elaborate on the implications of this later, when we discuss the continuous-state version of our model. For now, we assume that γ can take only three values.

First, γ can be γ_1 , which is such that $z_1 = 0$. This characterizes a relatively healthy consumer with zero health care demand. Second, γ can be $\gamma_2 > \gamma_1$, with $\gamma_2 > bt(1 - \beta(y - p)) > 0$. This characterizes a patient in need of health care and for whom co-insurance payments are strictly positive, but below the maximum m . For this patient, higher spending implies higher co-insurance payments.

Thirdly, γ may be $\gamma_3 > \gamma_2$ which implies that health care demand is sufficiently large to make the consumer pay the maximum of co-insurance payments m . The exogenous probabilities associated with the three cases are denoted as $\pi_1 \geq 0$, $\pi_2 > 0$ and $\pi_3 \geq 0$.

Health care demand takes the following form in the three cases:

$$z_1 = 0$$

$$z_2 = \frac{\gamma_2 - bt(1 - \beta(y - p))}{\delta + \beta(bt)^2} \quad (3)$$

$$z_3 = \frac{\gamma_3}{\delta}$$

An important feature is that demand is price-elastic only in the second state. Demand in states 1 and 3 is inelastic, albeit for different reasons. Using equations (2) and (1), it is straightforward to derive also expressions for non-medical consumption and utility in the three states distinguished, c_i and u_i , $i = 1, 2, 3$.

We assume that the health insurance industry is perfectly competitive and features zero administration costs. This is fully equivalent to the assumption of a national insurer, given that adverse selection does not play any role in our paper – due to the assumption that agents are identical before the realization of shocks. Hence, health insurance premiums equal health spending minus co-insurance payments:

$$p = \pi_2(1 - b)tz_2 + \pi_3(tz_3 - m) \quad (4)$$

2.2 The optimal co-insurance rate

Before the occurrence of health shocks, the representative health insurer chooses the co-insurance rate and co-insurance maximum that maximizes expected utility:

$$E(U) = \pi_1 u_1 + \pi_2 u_2 + \pi_3 u_3 \quad (5)$$

Appendix B derives that the co-insurance rate b^* that maximizes (5) can be written as

$$b^* = \frac{\Phi \varepsilon}{(1-\Phi) + \Phi \varepsilon} \quad (6)$$

where Φ is a help variable, defined as $(\pi_1 \mu_1 + \pi_2 \mu_2 + \pi_3 \mu_3) / \mu_2$, and μ_i

$i = 1, 2, 3$ are the marginal utilities of income in the three states, *i.e.*

$$\mu_1 = (1 - \beta(y - p)), \quad \mu_2 = (1 - \beta(y - p - btz_2)) \quad \text{and} \quad \mu_3 = (1 - \beta(y - p - m)).$$

Furthermore, ε is the elasticity of health care demand with respect to its price in state 2, which we will define in absolute terms throughout the paper.

Equation (6) allows us to derive the result expressed in proposition 1.

Proposition 1: If $\Phi < 1$, $b^* < 1$;
 if $\Phi = 1$, $b^* = 1$;
 if $\Phi > 1$, $b^* > 1$.

Proposition 1 states that, in general, with $\beta \geq 0$ and $\pi_1, \pi_2, \pi_3 \geq 0$, the optimal co-insurance rate can be lower than, equal to, or higher than unity. Three cases are of special interest. It is useful to start with the case $\beta = 0$ where the marginal utility of income is constant across the three states and $\Phi = 1$. So there

is no need for risk sharing. However, it is efficient to eliminate moral hazard.

Hence, the optimal co-insurance rate equals unity.

Next, assume $\beta > 0$ and $\pi_1 > 0, \pi_2 > 0, \pi_3 = 0$. In this case, $\Phi < 1$. Co-insurance implies a transfer from the state of bad health, state 2, to the state of perfect health, state 1. As the marginal utility of income is higher in the former state, this transfer in itself reduces expected utility. The optimal co-insurance rate is now less than unity, balancing the welfare loss from a lack of risk sharing with the welfare gain from a reduction of moral hazard.

The third case is exactly the opposite of the second one. Assume $\beta > 0$ and $\pi_1 = 0, \pi_2 > 0, \pi_3 > 0$. In this case, $\Phi > 1$. A higher co-insurance rate now implies higher transfers from the state of bad health, state 2, to the state of very bad health, state 3. As the marginal utility of income is highest in the latter state, this increase in transfers raises expected utility. The optimal co-insurance rate will now exceed unity, balancing the welfare gain from risk sharing with the welfare loss from an underconsumption of medical care. Obviously, in the real world we do not see schemes that feature co-insurance rates higher than unity. A rate higher than unity would run counter to the whole idea of insurance. If we impose therefore the restriction that the co-insurance rate cannot exceed unity, the third case becomes equivalent to the first case.

Anticipating the result derived below that an optimal insurance scheme features a finite maximum, the second and third case thus illustrate that it may be optimal to have a deductible. This confirms the result of Drèze and

Schokkaert (2013) that Arrow (1963)'s theorem of the deductible can also hold true in a model with ex post moral hazard.³

Equation (6) also tells us something about the relationship between the optimal co-insurance rate and the price elasticity of health care demand, if we assume that Φ can be treated as a constant. We will argue below that this is a useful assumption, but first present our result as proposition 2.

Proposition 2: If $\Phi < 1$, $\frac{\partial b^*}{\partial \varepsilon} > 0$;
 if $\Phi = 1$, $\frac{\partial b^*}{\partial \varepsilon} = 0$;
 if $\Phi > 1$, $\frac{\partial b^*}{\partial \varepsilon} < 0$

We distinguish the same three cases as before. In case $\beta = 0$, the optimal co-insurance rate equals unity, irrespective the value of the price elasticity of demand, which explains that $\partial b^* / \partial \varepsilon = 0$ if $\Phi = 1$. The second case with $\beta > 0$ and $\pi_3 = 0$ is more interesting. Then, the optimal co-insurance rate balances risk sharing and moral hazard. As has been derived before (Pauly 1968; Zeckhauser 1970; Besley 1988; Goldman and Philipson 2007; Zweifel, Breyer and Kifmann 2009; McGuire 2012)⁴, more elastic demand increases the optimal co-insurance rate in the direction of unity: $\partial b^* / \partial \varepsilon > 0$ if $\Phi < 1$. The third case

³ Gollier (2000) shows that Arrow's deductibility theorem can even be derived without assuming expected utility.

⁴ Einav, Finkelstein and Polyakova (2016) find that this theoretical result also holds true in practice: coinsurance rates and price elasticities are positively correlated in the case of US Medicare Part D, i.e. insurance of prescription drugs in private insurance plans.

($\beta > 0, \pi_1 = 0$) is again opposite to the second one. But if we impose the institutional requirement that the co-insurance rate cannot exceed unity, the third and first case again coincide.

As mentioned above, proposition 2 assumes that Φ is a constant. This conflicts with the definition of Φ as a combination of three marginal utilities of income. However, we argue that realistic variations in Φ will not destroy proposition 2 for two reasons: btz_2 and m are much smaller than $y - p$ and β is very small.

As regards the former argument, one may argue that the marginal utilities of income in state 1, 2 and 3 will be close to each other. Indeed, disposable income $y - p$ is generally much higher than health care spending below the cap, btz_2 , and also much higher than the co-payment maximum, m . Hence, Φ must be close to a constant. This holds true even for values of m close to $y - p$, as the probability of state 3 is small relative to those of states 1 and 2.

As regards the latter argument, observe that in the polar case case $\beta = 0$, Φ is a constant. Hence, on the basis of a continuity argument, Φ will vary little for values of β close to zero. In model applications like the one in section 4 of this paper, β is very small (indeed, this must be the case as large values would violate the inequality condition in equation (1)). Furthermore, simulations made with a numerical version of the model in this section confirm that the variation in Φ becomes smaller, the lower is the value of β .

Additionally, we have investigated the variability of Φ using a numerical version of the simple model in this section. These simulations (not included for

brevity) confirm that variations in the parameter δ (equation (3)) induce variations in ε and Φ in the same direction, just as stated in proposition 2.

2.3 The optimal co-insurance maximum

As the price elasticity of health care demand relates negatively to health care spending, imposing a cap makes health care demand below that cap more price-elastic. Following proposition 2, we may state that if the insurance scheme features a deductible, introducing a cap on co-insurance will not affect the optimal co-insurance rate. But if the co-insurance rate is below one, the reform will increase it.

What is the optimal co-insurance maximum? In order to explore this, we derive the first-order condition that corresponds to the maximization of expected utility with respect to the co-insurance maximum.

Using m^* to denote the optimal co-insurance maximum, Appendix B derives the following expression for this maximum:

$$m^* = \left(\frac{\pi_2}{1 - \pi_3} \right) btz_2 \quad (7)$$

Although equation (7) is not a reduced-form equation, it clearly indicates that the optimal co-insurance maximum must be finite.⁵

Appendix B clarifies that two assumptions are required to derive equation (7): $\beta > 0$ and $\pi_1 > 0$. If we would instead assume that $\beta = 0$ would hold true, the co-insurance maximum would be undefined as insurance would have zero

⁵ This result differs from the one reached by Manning and Marquis (1996). Using simulations, these authors were unable to find a plausible maximum. We suspect – as they do – that their assumption of a constant price elasticity plays a crucial role since the assumption that the price elasticity is decreasing in health status is crucial in our case.

value. If, instead, we would assume that $\pi_1 = 0$, m^* would equal btz_2 , which violates our definition of state 2 as one in which the patient pays less than the co-insurance maximum.⁶

That the optimal co-insurance maximum is finite can be explained by describing a schedule which allows the co-insurance rate to be different for different values of health status γ . We hypothesize that this schedule would feature a co-insurance rate that is decreasing in γ . For we know that the optimal co-insurance rate is increasing in ε (proposition 2) and ε is decreasing in γ (see appendix A); moreover, Blomqvist (1997) has derived exactly this result. In this case the optimal co-insurance maximum thus balances the welfare loss from a too high rate just below the maximum against the welfare loss from a too low rate (namely zero) just above the maximum. This argument implies that the higher the optimal co-insurance rate, the higher will be the optimal co-insurance maximum. This is indeed what equation (7) suggests. One should note that this only holds true if the price elasticity of health care demand is below one in absolute terms. What we know from the literature is that this is a quite innocent assumption as price elasticities are in the order of 10 to 20 percent.

3 A graphical illustration of the introduction of a cap

Before moving to a continuous-state version of the model it is useful to present a graphical analysis of the likely effects of the introduction of a cap on co-

⁶ This result implies that state 3 that we defined as $\beta > 0$ and $\pi_1 = 0, \pi_2 > 0, \pi_3 > 0$ in our discussion of propositions 1 and 2 is incompatible with the model. This does not invalidate our discussion or any of the propositions as these do not rely on specific values for π_1 (or π_2 or π_3).

insurance payments. Therefore, let us take a look at figure 1. This assumes a continuum of states and medical products. This figure displays hypothetical effects, namely how co-insurance payments would relate to the health status parameter γ if the health insurance premium and the marginal utility of income were constant. Here, the line 0AB represents co-insurance payments in the case without a cap. The line 0DAC represents co-insurance payments in the case with a cap: it puts a cap of m on co-insurance payments and features a higher co-insurance rate (the line 0DE is steeper than the line 0AB). The line 0FGH measures the difference between 0DAC and 0AB. The figure shows that upon the introduction of a cap consumers with γ higher than γ_G would gain and those with γ lower than γ_G would lose. Further, consumers with γ close to the level at which the cap starts to apply, γ_F , would be worst off.

In general, the welfare effects of the introduction of a cap do not coincide with the effects upon co-insurance payments. One reason is that the reform may change insurance premiums; another is that the marginal utility of income is not a constant. However, as verified by the simulations below, figure 1 gives a good characterization of the shape of the welfare effects of a co-insurance cap.

INSERT FIGURE 1 ABOUT HERE

4 Optimal co-insurance in a continuous-state world

4.1 Extensions of the model

This section extends the stylized model of section 2. In this more realistic model it is no longer possible to prove propositions 1 and 2. But, as we will see, simulations with this extended model indicate that also in this case both propositions still hold, and that the optimal co-insurance rate is below unity. Furthermore, the model enables to get an idea of the effects of the introduction of a cap on patient welfare.

So, rather than assigning a unique value for the parameter γ to each of the three states distinguished in section 2, we now assume that γ is a stochastic parameter with continuous distribution function $G(\cdot)$. Consequently, the probabilities π_i of specific co-insurance regimes (*i.e.* zero spending, spending lower than m , spending higher than m) are no longer fixed, but endogenous.

Next, we extend the model by introducing two distinct patient groups; patients that incur costs of outpatient care only (the O-group) and patients who are also admitted to a hospital (A-group). We assume that the parameters of the distribution function of the health care parameter γ differ between the two groups.

It is important to note that *ex ante* all patients are equal. Each period patients receive two – uncorrelated – negative health shocks. The first shock indicates what kind of care the patient needs to improve his or her health status. Some of them will need no care at all. We name this the N-group (No care). Others need outpatient care only (the O-group), or use all care (A): both outpatient and inpatient care. The second shock determines a particular value for γ , which

indicates, as before, the size of the intervention that is required to restore the patient's health status.

As regards $G(\cdot)$, the distribution function of γ , we assume the following. We write γ as $(\gamma - \gamma_{\min}) + \gamma_{\min}$, where γ_{\min} indicates a non-negative nonstochastic parameter that differs between the two groups⁷. $\gamma - \gamma_{\min}$ is then assumed to be lognormally distributed. The parameters of this distribution differ between the O-group and the A-group. The value for the parameter γ_{\min} in the two groups follows from calibrating the model to statistics for the consumption of O-care and A-care.

In deciding about the co-insurance scheme, we assume that the government does not distinguish between O-care and A-care. Alternatively, the government may differentiate the co-insurance scheme between the two types of care. However, in that case, it would have been more difficult to defend our assumption that patients cannot decide which type of care to receive.

4.2 Consumers

In the cases of O-care and A-care, consumers face an optimization problem like in section 2: they decide on their demand for health care given the value of γ . But now they do not know a priori which of the 3 cases (see equation (3)) will apply; that follows from their utility maximization. The budget constraint is kinked: if health care costs exceed the maximum of co-payments m , the out-of-

⁷ One may think of fixed costs, needed to refer patients to outpatient or inpatient care.

pocket price drops to zero. Following the optimization procedure spelled out in Westerhout and Folmer (2013), one can derive the following expressions for health care demand (where we leave out the subscripts O and A for brevity):

$$\begin{aligned}
 z = 0 & \qquad \qquad \qquad \gamma_{\min} \leq \gamma \leq \gamma_0 \\
 z = \frac{\gamma - bt(1 - \beta(y - p))}{\delta + \beta(bt)^2} & \qquad \gamma_0 \leq \gamma \leq \gamma_1 \\
 z = \frac{\gamma}{\delta} & \qquad \qquad \qquad \gamma_1 \leq \gamma
 \end{aligned} \tag{9}$$

The corresponding expressions for the endogenous boundary values γ_0 and γ_1

$$\begin{aligned}
 \gamma_0 &= bt(1 - \beta(y - p)) \\
 \text{and} & \\
 \gamma_1 &= \frac{-\delta(1 - \beta(y - p))}{\beta(bt)} + \frac{\delta}{\beta(bt)^2} \sqrt{\Omega}
 \end{aligned} \tag{10}$$

with

$$\begin{aligned}
 \Omega &= (bt)^2 (1 - \beta(y - p))^2 \\
 &+ 2(bt)^2 \beta \left[\frac{1}{2} \beta m^2 + m(1 - \beta(y - p)) + \frac{(bt)^2}{\delta} (1 - \beta(y - p - m))^2 \right]
 \end{aligned} \tag{11}$$

The condition $\gamma_1 > \gamma_0$ in equation (9) applies if the co-insurance maximum m exceeds some small minimum value. We do not explore the case in which the co-insurance maximum is below this minimum (this case does not occur in any of our simulations). Further, we make two observations. First, equation (9)

implicitly assumes that $\gamma_{\min} < \gamma_0$. In case $\gamma_{\min} \geq \gamma_0$, equation (9) applies as well, except for the corner solution $z = 0$. Second, in case of an unbounded co-insurance scheme, γ_1 is infinite and hence the third segment of equations (9) and (10) does not apply.

Expressions for co-insurance payments, non-medical consumption and indirect utility can be easily obtained by combining the expressions for health care demand in equation (9) with the budget constraint (2) and the direct utility function (1). For brevity, we do not report them here.

4.3 Health insurers

For both O-care and A-care we integrate the levels of consumer utility for all values of γ to arrive at expressions for expected utility. Let $G_i(\cdot)$ denote the distribution function of $\gamma_i - \gamma_{i,\min}$ ($i = O, A$), then:

$$\begin{aligned}
E(v_i) &= G_i(\gamma_{i,0}) E(v_i | \gamma_i \leq \gamma_{i,0}) \\
&+ (G_i(\gamma_{i,1}) - G_i(\gamma_{i,0})) E(v_i | \gamma_{i,0} \leq \gamma_i \leq \gamma_{i,1}) \\
&+ (1 - G_i(\gamma_{i,1})) E(v_i | \gamma_i \geq \gamma_{i,1}) \quad i = O, A
\end{aligned} \tag{12}$$

Expressions for the conditional expectation variables $E(v|\cdot)$ in equation (12) are based upon the expression for indirect utility that corresponds to equation (9).

Expected utility is a weighted average of $E(v_O)$, $E(v_A)$ and $E(v_N)$,

$$V = \pi_N v_N + (1 - \pi_N) \pi_O E(v_O) + (1 - \pi_N) (1 - \pi_O) E(v_A) \tag{13}$$

where v_N , utility in case of zero care, equals $(y - p) - \frac{1}{2}\beta(y - p)^2$.

The expression for the health insurance premium p completes the model in this section. As before, this premium equals the difference between the aggregates of spending and co-insurance payments.

4.4 Calibration

We use observed relative sizes of the N-, O- and A-groups, and total copayments from a sample for the Netherlands. We apply estimated means and standard deviations of spending on outpatient and total care to the parameters of the distribution parameters of γ_k , $k = O, A$. The average coefficient of relative risk aversion is set at 2 (Garber and Phelps 1997). Price elasticities of O-care and A-care are -0.079 and -0.011 (corresponding to Van Vliet 2001). See Appendix C for more details.

The estimated price elasticities are low as compared to the price elasticities from the RAND Health Insurance Experiment (Newhouse and the Insurance Experiment Group, 1993). Below, we will explore the sensitivity of our results with respect to these price elasticities.

4.5 Numerical simulations

To estimate the optimal co-insurance rate for the unbounded scheme, we apply a grid search procedure in which we evaluate co-insurance rates in the range from 0% to 100% in steps of 5 percentage points. Table 1 shows that the optimal co-insurance rate is 30%.

We have performed a sensitivity analysis to explore how robust this result is (not reported for brevity). In particular, we have varied the degree of risk aversion, the price elasticities of demand, the standard deviations of the distributions of health status and the weight of outpatient care in total health care. The results display quite some variation: the co-insurance rate varies between 10% and 55%. On this point, our results are in line with Manning and Marquis (1996) who report a range for the optimal co-insurance rate from 25 to 50%. Also, these results are in line with proposition 1: the optimal co-insurance rate will be below one if consumers are risk-averse, and the probability of consumers having to pay the maximum of co-insurance payments is zero.

INSERT TABLE 1 ABOUT HERE

Table 1 displays the effects of introducing a co-insurance maximum. Here again, we apply a grid procedure. Now we also vary the co-insurance maximum. For each value of the co-insurance rate, we explore the range from 50 to 5000 euro. This globally locates the maximum. Then we repeat the procedure choosing values for m around the maximum from the first step, and so on. The reform results in a co-insurance maximum of about 3250 euro and a co-insurance rate of 55%. Corresponding to the introduction of the cap and in line with our analysis in section 2, the co-insurance rate thus increases (from 30 to 55%). Co-insurance payments rise by about 40%, whereas aggregate health spending and health insurance premiums fall.

Average co-insurance payments for O-care and A-care increase as well. This may be surprising as the direct impact of putting a cap on co-insurance is a reduction in co-insurance payments. It appears that a sizeable fraction of the consumers of A-care spends less than the imposed maximum. These people face an increase of co-insurance payments on account of a higher co-insurance rate. This effect dominates the direct effect of the cap.

To obtain a meaningful measure of the welfare gain, we calculate the compensating variation, *i.e.* the income loss that the average insured person would be willing to accept to prevent him from being moved to an alternative scheme. Formally, the compensating variation of the bounded scheme, \tilde{y}^B , is defined by the equality $V(b^B, m^B, y - \tilde{y}^B) = V(b^U, m^U, y)$, where V denotes expected utility and b^U and m^U are the co-insurance rate and maximum of the unbounded scheme and, similarly, b^B and m^B are the counterparts of the bounded scheme. We apply this definition at the insured population, to the O and A groups and to the individual level. We find that the compensating variation at the aggregate level is very modest: 9 euro per insured only. This aggregate welfare gain hides substantial transfers between groups of consumers. The consumers of A-care lose 201 euro, whereas the O-care patients gain 29 euro. Both groups share the gain from lower insurance premiums, but the consumers of A-care suffer most from the increase in the co-insurance rate.

The distribution of welfare effects indicates a lot of heterogeneity within the O and A groups. In both groups, the patient whose health care expenditures are close to the co-payment maximum is hurt most. The maximum individual loss

for a consumer in the O group amounts to 1,343 euro, whereas in the A group this maximum loss is 1,317 euro.

The increase of the co-insurance rate due to the introduction of the cap generates this uneven redistributive effects. If the co-insurance rate had been fixed to its pre-reform value, the maximum loss for a consumer in the O group would have dropped to 30 euro only, which contrasts sharply with the amount of 1,343 euro calculated earlier.

These large numbers for the least well-off consumers are confirmed in our sensitivity analysis. Table 2 shows the results for eight alternative simulations. In most cases the maximum losses are very large, both for O-care and A-care.

INSERT TABLE 2 ABOUT HERE

5 Concluding remarks

Any analysis needs a lot of assumptions and our paper is no exception: we have assumed linear-quadratic preferences, no role for suppliers of health services and full information on part of consumers about their health status after health shocks have realized. We could have made alternative assumptions. However, as long as alternative preferences imply that health care consumption falls when its price increases, as long as health care demand has some effect upon consumption and as long as consumers have some information about their health status, our result that the distribution of welfare effects of the introduction of a co-insurance cap is very uneven, will probably survive,

provided that two assumptions are maintained. First, health care demand becomes less price-elastic when health status deteriorates and second, insurers increase the co-insurance rate when demand becomes more price-elastic.

The authors declare that there are no potential conflicts of interest.

This paper (or parts of it) has not been published before and has not been submitted to a different journal.

There is no ethical background to this study.

References

- Arrow, K.J., 1963. Uncertainty and the Welfare Economics of Medical Care, *American Economic Review* **53**, 941-973.
- Besley, T.J., 1988. Optimal Reimbursement Health Insurance and the Theory of Ramsey Taxation, *Journal of Health Economics* **7**, 321-336.
- Blomqvist, A., 1997. Optimal Non-Linear Health Insurance, *Journal of Health Economics* **16**, 303-21.
- Buchmueller, T., 2006. Price and the health plan choices of retirees, *Journal of Health Economics* **25**, 81-101.
- Chandra, A., J. Gruber and R. McKnight, 2010. Patient Cost-Sharing and Hospitalization Offsets in the Elderly, *American Economic Review* **100**, and 193-213.
- Chandra, A., J. Gruber and R. McKnight, 2014. The Impact of Patient Cost-Sharing on low-income populations: Evidence from Massachusetts, *Journal of Health Economics* **33**, 57-66.
- Drèze, J.H. and E. Schokkaert, 2013. Arrow's theorem of the deductible: moral hazard and stop-loss in health insurance, *Journal of Risk and Uncertainty* **47**, 147-163.
- Eggleston, K., 2000. Risk selection and optimal Health Insurance-Provider Payment Systems, *The Journal of Risk and Insurance* **67**(2), 173-196.
- Einav, L., A. Finkelstein and M. Polyakova, 2016. Private Provision of Social Insurance: Drug-Specific Price Elasticities and Cost Sharing in Medicare Part D, NBER Working Paper 22277.

- Ellis, R.P., 1986. Rational Behavior in the Presence of Coverage Ceilings and Deductibles, *RAND Journal of Economics* **17**, 158-75.
- Ellis, R.P. and Th.G.McGuire, 1990. Optimal Payment Systems for Health Services, *Journal of Health Economics* **9**, 375-96.
- Engelhardt, G.V. and J. Gruber, 2011. Medicare Part D and the Financial Protection of the Elderly, *American Economic Journal: Economic Policy* **3**, 77-102.
- Finkelstein, A., E.F.P. Luttmer and M.J. Notowidigdo, 2013. What Good is Wealth Without Health? The Effect of Health on the Marginal Utility of Consumption, *Journal of the European Economic Association* **11**, 221-258.
- Finkelstein, A. and R. McKnight, 2008. What did Medicare do? The initial impact of Medicare on mortality and out of pocket medical spending, *Journal of Public Economics* **92**, 1644-1668.
- Garber, A.M. and C.E. Phelps, 1997. Economic foundations of cost-effectiveness analysis, *Journal of Health Economics* **16**, 1-31.
- Glazer, J. and T.G. McGuire, 2012. A Welfare Measure of “Offset Effects” in Health Insurance, *Journal of Public Economics* **96**, 520-23.
- Goldman, D. and T.J. Philipson, 2007. Integrated Insurance Design in the Presence of Multiple Medical Technologies, *American Economic Review* **97**, 427-432.
- Gollier, C., 2000. Optimal insurance design: What can we do with and without expected utility?, in G. Dionne (ed.), *Handbook of Insurance*, Kluwer Academic Publishers, London, 97-116.

- Keeler, E.B., J.P. Newhouse and C.E. Phelps, 1977. Deductibles and the Demand for Medical Care Services: The Theory of a Consumer Facing a Variable Price Schedule under Uncertainty, *Econometrica* **45**, 641-56.
- Kowalski, A.E., 2012. Estimating the Tradeoff between Risk Protection and Moral Hazard with a Nonlinear Budget Set Model of Health Insurance, NBER working paper 18018.
- Manning, W.G. and M.S. Marquis, 1996. Health Insurance: the Tradeoff between Risk Pooling and Moral Hazard, *Journal of Health Economics* **15**, 609-39.
- Ma, C.A. and Th.G. McGuire, 1997. Optimal Health Insurance and Provider Payment, *American Economic Review* **87**, 685-704.
- McGuire, Th.G., 2012. *Demand for Health Insurance*, in M.V. Pauly, Th.G. McGuire and P.P. Barros (eds.), *Handbook of Health Economics*, vol. 2, Elsevier, 317-396.
- Newhouse, J.P. and the Insurance Experiment Group, 1993. *Free for All? Lessons from the Health Insurance Experiment*, Harvard University Press, Cambridge.
- Newhouse, J.P., 2006. Reconsidering the moral hazard-risk avoidance trade off, *Journal of Health Economics* **25**, 1005-14.
- Nyman, J.A., 1999. The value of health insurance: the access motive, *Journal of Health Economics* **18**, 141-152.
- OECD, 2015. Health policies and data, Cost-sharing requirements and exemptions of co-payments for different population groups.

www.oecd.org/els/health-systems/Coverage-cost-sharing-and-exemptions.xlsx.

Pauly, M.V., 1968. The economics of moral hazard: comment, *American*

Economic Review **58**, 531-537.

Pauly, M.V. and F.E. Blavin, 2008. Moral hazard in insurance, value based cost-sharing and the benefits of blissful ignoring, *Journal of Health*

Economics **27**, 1407-17.

Phelps, C.E. and J.P. Newhouse, 1974. Coinsurance, the Price of Time, and the Demand for Medical Services, *Review of Economics and Statistics* **56**, 334-

42.

Statistics Netherlands (2006), Working paper *Health Accounts 1998-2004*, The Hague.

Strombom, B.A., T.C. Buchmueller and P.J. Feldstein, 2002. Switching costs, price sensitivity and health plan choice, *Journal of Health Economics* **21**, 89-

116.

Van Vliet, R.C.J.A., 2001. Effects of Price and Deductibles on Medical Care Demand, Estimated from Survey Data, *Applied Economics* **33**, 1515-24.

Van Vliet, R.C.J.A. and H.G. Van der Burg, 1996. *Distribution Functions of Costs of Health Care Services*, iBMG, Erasmus University Rotterdam (in

Dutch).

Wedig, G.J., 1988. Health Status and the Demand for Health – Results on Price Elasticities, *Journal of Health Economics* **7**, 151-63.

Westerhout, E.W.M.T. and C.E. Folmer, 2013. Why it may hurt to be insured: the effects of capping coinsurance payments, CPB Discussion Paper 239.

Zeckhauser, R., 1970. Medical insurance: a case study of the tradeoff between risk spreading and appropriate incentives, *Journal of Economic Theory* **2**, 10-26.

Zhou, Z., S. Yanfang, G. Jianmin, X. Ling and Z. Yaoguang, 2011. New estimates of elasticity of demand for healthcare in rural China, *Health Policy* **103**, 255– 65.

Zweifel, P., F. Breyer and M. Kifmann, 2009. *Health Economics*, Springer, Dordrecht.

Figure 1: Co-insurance payments in the linear scheme and the scheme with a cap as a function of the health status parameter γ

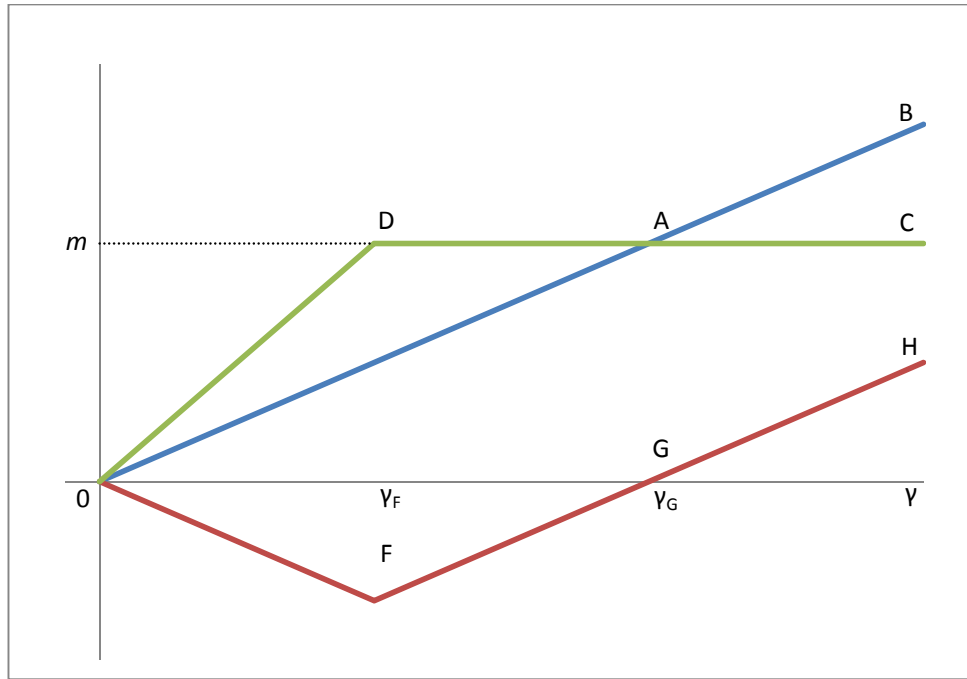


Table 1 Effects of reforming the co-insurance scheme

	<i>U</i>	<i>B</i>	$(B-U)/U$ (%)
Optimal co-insurance rate (%)	30	55	83.33
Optimal co-insurance maximum	50,000	3,250	93.50
Aggregate health care consumption	15.82	15.26	-3.54
- of which: O group	15.16	14.55	-4.02
- of which: A group	23.44	23.39	-0.21
Health care spending	942.27	925.07	-1.83
Co-insurance payments	282.67	395.50	39.92
Co-insurance payments, O group	132.55	224.42	69.31
Co-insurance payments, A group	2,009.03	2,362.99	17.62
Insurance premium	659.60	529.57	-19.71

U: The optimal unbounded co-insurance scheme

B: The optimal bounded co-insurance scheme

Table 2 Sensitivity analysis of the welfare loss of the least well-off patients

	$\hat{b}(\%)$		\hat{m}	\tilde{y}^U	$\max(\tilde{y}_O^U)$	$\max(\tilde{y}_A^U)$
	<i>U</i>	<i>B</i>				
Benchmark (BM)	30	55	3,250	9	-1343	-1317
$\beta = 0.50\beta_{BM}$	55	70	7,100	8	-1450	-1400
$\beta = 1.50\beta_{BM}$	20	85	1,100	11	-690	-661
$\delta_i = 0.50\delta_{i,BM}$	45	65	5,500	12	-1586	-719
$\delta_i = 1.50\delta_{i,BM}$	20	85	1,150	8	-1513	-639
$\sigma_i = 0.50\sigma_{i,BM}$	45	45	7,500	0	-4	-4
$\sigma_i = 1.50\sigma_{i,BM}$	10	60	2,450	22	-1796	-1786
$\pi_O = 0.95\pi_{O,BM}$	30	45	3,500	10	-1510	-1489
$\pi_O = 1.05\pi_{O,BM}$	40	90	1,100	10	-1343	-1317

$i = O, A$; σ_i is the variance of $\ln(\gamma_i - \gamma_{i,\min})$; π_O is the relative share of the O group in the population with positive health care costs.

U: The optimal unbounded co-insurance scheme

B: The optimal bounded co-insurance scheme

Appendix A: Properties of the price elasticity of health care demand

1 *The price elasticity of health care demand is a decreasing function of γ*

Recall our expression for health care demand (equation (3)):

$$z = \frac{\gamma_2 - bt(1 - \beta(y - p))}{\delta + \beta(bt)^2}$$

Differentiate z with respect to bt and divide the result by $z / (bt)$ to find an expression

for the price elasticity of health care demand $\varepsilon_p \equiv (\partial z / \partial(bt)) / (z / (bt))$:

$$\varepsilon_p = \frac{-2\gamma_2\beta(bt)^2 - (1 - \beta(y - p))bt(\delta - \beta(bt)^2)}{(\gamma_2 - (1 - \beta(y - p))bt)(\delta + \beta(bt)^2)} \quad (\text{A1})$$

Equation (A1) uses the partial derivative $\partial z / \partial(bt)$ to define the price elasticity. The elasticity ε used in the main text is based on the total derivative $dz / d(bt)$. The latter derivative includes also the effect of a change in out-of-pocket price upon health care demand through a change in insurance premiums. Numerically, the difference between the two elasticities is negligibly small.

To find the role of the parameter γ_2 , we re-write equation (A1) as follows:

$$\varepsilon_p = \left\{ \frac{-2\beta(bt)^2}{\delta + \beta(bt)^2} \right\} - \left\{ \frac{(1 - \beta(y - p))bt}{\gamma_2 - (1 - \beta(y - p))bt} \right\} \quad (\text{A2})$$

From this expression, it can easily be derived that $\partial \varepsilon_p / \partial \gamma_2 > 0$. Hence, in absolute terms, the price elasticity of health care demand decreases upon an increase of the health status.

2 *The price elasticity of health care demand is a decreasing function of bt*

Take equation (A2). The effect of an increase in bt is to increase the value of both the first and the second term, which enter into the expression for ε_p with a minus sign.

Hence, upon an increase in bt , ε becomes more negative, *i.e.* $\partial\varepsilon_p / \partial(bt) < 0$.

3 The price elasticity of health care demand is a decreasing function of health spending tz

Write down the reduced-form expression for health spending tz :

$$tz = \frac{t(\gamma_2 - bt(1 - \beta(y - p)))}{\delta + \beta(bt)^2}$$

Form this expression, it follows that $\partial(tz) / \partial\gamma_2 > 0$.

Write the derivative of the price elasticity of health care demand with respect to health spending as follows:

$$\frac{\partial\varepsilon_p}{\partial(tz)} = \frac{\partial\varepsilon_p}{\partial\gamma_2} \bigg/ \frac{\partial(tz)}{\partial\gamma_2}$$

Given that $\partial\varepsilon_p / \partial(tz) > 0$ and that $\partial(tz) / \partial\gamma_2 > 0$, $\partial\varepsilon_p / \partial\gamma_2 > 0$. Hence, in absolute terms, the price elasticity of health care demand decreases upon an increase of health spending.

Appendix B: Derivation of the first-order conditions with respect to the coinsurance rate and the coinsurance maximum

We start to elaborate the expression for expected utility (equation (5) in the main text) as a function of z_2 and z_3 :

$$\begin{aligned}
 E(U) = & \pi_1 \left\{ (y-p) - \frac{1}{2} \beta (y-p)^2 \right\} \\
 & + \pi_2 \left\{ (y-p-btz_2) - \frac{1}{2} \beta (y-p-btz_2)^2 + \gamma_2 z_2 - \frac{1}{2} \delta z_2^2 \right\} \\
 & + \pi_3 \left\{ (y-p-m) - \frac{1}{2} \beta (y-p-m)^2 + \gamma_3 z_3 - \frac{1}{2} \delta z_3^2 \right\}
 \end{aligned} \tag{B1}$$

We use this equation to obtain the following version of the first-order condition with respect to the coinsurance rate:

$$\begin{aligned}
 \frac{dE(U)}{db} = & -\pi_1 \{1 - \beta(y-p)\} \frac{dp}{db} \\
 & -\pi_2 \{1 - \beta(y-p-btz_2)\} \frac{dp}{db} - \pi_2 \{1 - \beta(y-p-btz_2)\} tz_2 \\
 & -\pi_2 [(1 - \beta(y-p-btz_2))bt - \gamma_2 + \delta z_2] \frac{dz_2}{db} \\
 & \pi_3 \{1 - \beta(y-p-m)\} \frac{dp}{db} = 0
 \end{aligned} \tag{B2}$$

Collecting all dp/db terms on the RHS and making use of the envelope theorem so that the third line of equation (B2) vanishes, we rewrite the first-order condition in the following way,

$$-\Phi \frac{dp}{db} - \pi_2 tz_2 = 0 \tag{B3}$$

where the help variable Φ is defined as follows:

$$\Phi \equiv \frac{\pi_1(1-\beta(y-p))}{1-\beta(y-p-btz_2)} + \pi_2 + \frac{\pi_3(1-\beta(y-p-m))}{1-\beta(y-p-btz_2)} \quad (\text{B4})$$

Equation (B4) can be further simplified. Note that the insurance contribution p can be expressed as follows:

$$p = \pi_2(1-b)tz_2 + \pi_3(tz_3 - m) \quad (\text{B5})$$

Taking the derivative with respect to the coinsurance rate gives a convenient expression,

$$\frac{dp}{db} = -\pi_2tz_2 \left(1 + \frac{(1-b)}{b} \varepsilon \right) \quad (\text{B6})$$

where ε is defined as the (negative) of the price elasticity of health care demand in state 2, *i.e.* $\varepsilon \equiv -(dz_2 / dbt) / (z_2 / (bt))$.

Combining equations (B3) and (B6), dividing the result by π_2tz_2 and multiplying it by b , gives the expression for the optimal coinsurance rate in the main text:

$$b^* = \frac{\Phi \varepsilon}{1 + \Phi(\varepsilon - 1)} \quad (\text{B7})$$

To find the expression for the optimal coinsurance maximum, we differentiate the expression for expected utility (B1) with respect to m :

$$\begin{aligned} \frac{dE(U)}{dm} = & -\pi_3 \{1 - \beta(y - p - m)\} \\ & - \left[\pi_1 \{1 - \beta(y - p)\} + \pi_2 \{1 - \beta(y - p - btz_2)\} + \pi_3 \{1 - \beta(y - p - m)\} \right] \frac{dp}{dm} \end{aligned} \quad (\text{B8})$$

Equation (B5) shows that $dp / dm = -\pi_3$. Substitution of this into equation (B8) gives the following expression:

$$\beta m^* - \beta \left(\frac{\pi_2}{1 - \pi_3} \right) btz_2 = 0 \quad (\text{B9})$$

Equation (B9) shows that for $\beta = 0$ the optimal coinsurance maximum is undefined.

If $\beta > 0$ equation (7) in the main text results. Note that there is another regulatory condition that follows from the budget equation in the main text (equation 2): $m > btz_2$. Basically, this says that for the model to be applicable, coinsurance payments in the second state must be sufficiently lower than the coinsurance maximum. Combining this with the optimal value of m from equation (B9) we obtain that $\pi_2 / (1 - \pi_3) > 1$ or $\pi_1 > 0$.

Appendix C: Calibration of model parameters

The calibration upon data for the privately insured implies that in the sample insurance is complete, except for a deductible: the coinsurance rate equals 1. The maximum of coinsurance payments is chosen such that implied co-payments match the data. The probability of need for both inpatient and outpatient services, π_A , is measured by the ratio of the number of patients admitted to hospital to the total number of patients. We approximate this by the ratio of hospital admissions per patient.

For both O care and A care, we assume that the medical need parameter γ (in deviation from γ_{\min}) is lognormally distributed. For calibration of the two lognormal distributions, we use the estimates in Van Vliet and Van der Burg (1996).

Finally, the value of γ_{\min} is chosen such that the related health spending equals the sum of the costs of a one-day hospital admission, one consult of a medical specialist and one consult of a general practitioner. The latter is included as in The Netherlands it is required to visit a general practitioner before being allowed to consult a medical specialist. This amounts to 2000 euro (data obtained from Statistics Netherlands: <http://statline.cbs.nl/Statweb/>).

The remaining parameters cannot be directly computed from observed data. Therefore we use an iterative procedure to guarantee that the values of these unknown parameters are set such that the model meets observed data on a number of key variables. These observed key variables are:

1. Total health care demand for both groups (O and A)
2. The chosen value of the coefficient of relative risk aversion (CRRA: 2);
3. The price elasticity of demand for all outpatient services;
4. The insurance effect corresponding to all care services: the demand of a fully insured patient divided by the demand of an uninsured patient;
5. The observed probability of zero costs;
6. Total co-payments per patient.

Table C.1 summarizes the data used to calibrate the model. Table C.2 discusses economic features of the calibrated model. Table C.3 shows the allocation over different groups. It indicates that the number of consumers that ex post spend zero euros on health care is substantially higher than the probability of zero need: 2 percent. This is due to the fact that there is a large group of consumers that received a health shock that is so small that the benefits from health care consumption would be less than the costs involved.

Table C.1 Validation of model parameters: data

	Group O	Group A	Total
Probability of zero health care spending (%)			22.1
Probability of positive health care spending , A services (%)			8.0
Insurance effect A group		1.04	
Price elasticity of health care demand, O group	- 0.079		
CRRA, non-health products			2.0
Coefficient of variation health care spending	2.03	2.11	
Average demand health care services	14.5	24.9	
Real producer price health care services (euro)	24.4	239.4	
Real income per patient (euro)			35,321
Average co-payments per patient (euro)			164
Coinsurance rate (%)			100

Table C.2 Validation of model parameters: results

	Group O	Group A	Total
Probability of zero need: π_N (%)			2.0
Relative size of the O group: $(1 - \pi_N)\pi_O$ (%)			90.2
Relative size of the A group: $(1 - \pi_N)\pi_A$ (%)			8.0
Parameter β in utility function			$1.94 \cdot 10^{-5}$
Parameter δ in utility function: δ	3.0	27.0	
Expectation of $\log(\gamma - \gamma_{\min})$: μ	3.051	5.845	
Standard deviation of $\log(\gamma - \gamma_{\min})$: σ	1.278	0.896	
Average need per patient: $E(\gamma) / \delta$	15.9	25.0	
Minimum need per patient: γ_{\min} / δ	0.0	5.5	
Average co-payments per patient			159
Co-payment maximum	300	300	
Income elasticity health care consumption	0.16	0.0	0.15

Table C.3 Allocation of the population to budget segments

Patient group	total fraction	zero demand	decreasing budget segment	flat budget segment
No need (%)	2.1	2.1		
Positive need for health care, O group (%)	89.9	20.0	38.6	31.3
Positive need for health care, A group (%)	8.0	0.0	0.0	8.0