Why health insurance may hurt: the surprising effect of a cap on coinsurance payments

Ed Westerhout[a,b] and Kees Folmer[a,*]

[a] CPB Netherlands Bureau for Economic Policy Analysis, The Hague, The Netherlands

[b] University of Amsterdam, Department of Economics and Econometrics, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands.

*** Correspondence to: Kees Folmer, CPB Netherlands Bureau for Economic Policy Analysis, P.O. Box 80510, 2508 GM The Hague, The Netherlands; phone +31 70 3383397, fax: +31 70 3383350, e-mail: C.Folmer@cpb.nl.**

Abstract

Health insurance schemes usually apply cost sharing to curb moral hazard. It is common to limit this cost sharing, by applying a stop loss, for example. This can be motivated from an insurance perspective: without a cap, coinsurance payments might impose too much risk on consumers. This paper shows that introducing a cap on coinsurance payments has another effect and may actually *hurt* people with high medical costs. This is not due to moral hazard that comes along with the extra insurance. We show that introducing a cap makes health spending below the cap more price elastic. Health insurers that choose the coinsurance rate such as to optimally balance moral hazard and risk reduction, are then induced to raise the coinsurance rate.

Keywords: Moral Hazard, Caps, Coinsurance, Health Insurance Schemes, Welfare

JEL codes: D60, H21, I18

1 Introduction

Health care reform is high on the political agenda. The last few decades have witnessed a remarkable growth in health spending, due to, among others, demography, income growth and medical-technological progress (OECD 2011). As these trends are expected to last for some time, health spending growth is likely to continue (Hall and Jones 2007), (Chandra *et al.* 2013), (Westerhout 2014).

One way to obtain lower health care spending is to increase the role of cost sharing or coinsurance. In general, cost sharing induces patients to reduce their consumption of health care services. If cost sharing reduces consumption of which the social benefits are lower than the costs, it will increase welfare. Cost sharing increases financial uncertainty as well, however. Hence, optimal insurance will typically be partial insurance, striking a balance between the loss from moral hazard and the gain from risk sharing (Pauly 1968, Zeckhauser 1970).

There is no obvious reason to expect that the optimal cost sharing scheme is linear. Blomqvist (1997) explores analytically the properties of optimal nonlinear health insurance schemes. He shows that the optimal coinsurance rate depends on individual health status. In particular, the optimal coinsurance rate is decreasing in the severity of the health shock (*i.e.*, those hit by a more severe health shock should face a smaller marginal coinsurance rate).

There is more to say on optimal coinsurance schemes. Ellis and McGuire (1990) point to the role of supply. Generally, consumption will be determined by both demand and supply. This holds especially in health care markets in which the supplier is usually better informed than the patient. Newhouse (2006) and Chandra *et al*. (2010) point to the role of offset effects. If insurance reduces spending on a typical service like that of a general practitioner, but, because of this lower spending, increases it on other services like hospital services or pharmaceuticals, this will affect the moral hazard effect of insurance. Glazer and McGuire (2012) discusses the welfare impact of these offset effects and the role of insurance. Furthermore, lack of will power on part of the consumers of medical services – as stressed by behavioral economists, may be an argument against copayments (Newhouse 2006). Related, information imperfections on part of consumers may affect the case for coinsurance as well and may call for some sort of value-based insurance design (Pauly and Blavin 2008).

The goal of this paper is to analyze the distributional and welfare effects of a cap on coinsurance payments. A cap on coinsurance payments can be motivated from an insurance perspective: without a cap, coinsurance payments might impose too much risk on consumers. Caps are also common in health insurance schemes. 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 copayments can also be found in other countries, such as Austria, Finland, Iceland, Ireland, Norway, Sweden and Switzerland (OECD 2015). In addition, most health insurance schemes can be regarded as piecewise linear, which is also the type of insurance scheme that this paper will focus on.

In order to explore the welfare effects of a coinsurance maximum, the paper explores the effect of a reform that replaces an unbounded linear insurance scheme with a bounded scheme, *i.e.* a scheme that puts an upper bound on coinsurance payments. In both cases, the health insurer chooses the coinsurance rate such as to maximize welfare from an ex ante point of view. In the latter case, he also chooses the coinsurance maximum optimally.

The reform increases welfare, which is what one would expect when the set of policy parameters is enlarged. That the reform affects people that vary in their health status differently, is also quite intuitive. What is striking is that those who loose from the reform are people in relatively poor health. Indeed, those with medical spending at the level at which the maximum of coinsurance payments starts to apply, are hurt the most. The reason is, as this paper will show, that a cap on coinsurance payments induces health insurers to increase the coinsurance rate.

The increase in the coinsurance rate makes the distribution of welfare effects more uneven. Particularly, without an increase in the coinsurance rate, people with copayments that are initially below the maximum would all be hurt to the same extent (by the increase in the insurance premium). In our simulations, the distribution of welfare losses for this group is inverse V-shaped, however, which is fully explained by the increase of the coinsurance rate.

To a very large extent, the increase of the coinsurance rate determines the pattern of welfare gains and losses over the patient population. Hence, this result survives (large) changes in the set of parameter values, although the quantitative effects of the policy reform may differ considerably between different simulations.

There are several contributions to the literature that relate to our work, *e.g.* Feldstein (1973) and Feldman and Dowd (1991). Their assumption that the budget constraints of households are linear is not fruitful in our case, however. In order to properly assess the effects of a coinsurance maximum, it is necessary to adopt a nonlinear approach. Keeler *et al.* (1977), Ellis (1986), Manning and Marquis (1996) and Kowalski (2012) did account for the nonlinearity of the consumer's budget constraint, but did not explore the implications for the optimal cost sharing scheme. Also related to our paper are the papers by Finkelstein and McKnight (2008), Feldstein and Gruber (1995) and Engelhardt and Gruber (2011). These papers assess the welfare implications of moral hazard and risk reduction separately, using different models. In our view, this approach is too rough when the goal is to find optimal schemes. Therefore, we adopt an integral approach which assesses simultaneously the changes in moral hazard and risk reduction and the associated welfare effects.

The structure of our paper is as follows. Section 2 adopts a stylized three-state model to set out the logic of the argument. It uses this model to derive an analytical expression for the relationship between the optimal coinsurance rate and the price elasticity of health care demand. Section 3 extends the model into a continuous-state version and explains how to deal with the endogeneity of the price of health care services. Section 4 presents the results of numerical simulations with this extended model. Finally, section 5 contains concluding remarks.

2 How the price elasticity of health care demand affects the optimal coinsurance rate

In case of a cap on coinsurance payments health care demand is price-elastic – when spending is less than the cap - or price-inelastic – in case spending exceeds the cap. 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. Exploring the relation between the optimal coinsurance rate and the price elasticity of health care demand, we derive that the optimal coinsurance rate is increasing in the price elasticity of health care demand (in absolute terms). This result has been stated before (Zeckhauser 1970, Cutler and Zeckhauser 2000 and McGuire 2012), but not in case health care demand consists of the three categories that we distinguish.

We adopt a representative-agent approach. Hence, *ex ante* all consumers are alike. *Ex post* however, *i.e.* after the resolution of uncertainty, the population is heterogeneous. Heterogeneity steps in as people who are identical from an *ex ante* view encounter different shocks to their health. In other respects, we stay close to the standard model, that is, we abstain from modelling explicitly the role of suppliers, information imperfections and any offset effects in order to highlight the basic effects of a cap on copayments.

Ellis and McGuire (1990) and Eggleston (2000) use a quadratic form to describe the relation between utility and medical care. We do the same, primarily for two reasons. First, quadratic utility implies that the marginal utility of health care will become zero for some finite amount of health care. This allows one to describe the demand for health care also in case of a zero out-of-pocket price. This is an important characteristic, as zero out-of-pocket prices are quite common in the health care sector. Second, there are three types of empirical evidence that health care demand is less price-elastic for higher amounts of health care spending. First, Wedig (1988) has found that the price elasticity of health care demand is smaller the worse the health status of the patient. Second, Newhouse *et al.* (1993), Van Vliet (2001), Zhou *et al.* (2011) and Chandra *et al.* (2012) present evidence that the demand for expensive inpatient hospital services is less price-elastic than that for less expensive other medical services. Thirdly, Strombom *et al.*(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 demand by people who are older and who have been recently hospitalized or diagnosed with cancer. The quadratic specification is consistent with this type of evidence: it implies a price elasticity of health care demand that is decreasing in health status and spending (see Appendix A for a formal derivation).

Moreover, the quadratic form implies that the price elasticity of health care demand is increasing in the co-insurance rate (see Appendix A). This feature is backed by empirical evidence as well (Phelps and Newhouse 1974). The quadratic form has a drawback too. In particular, it implies that risk aversion is increasing in income. In our analysis, this argument has little weight since it does not account for income heterogeneity. In section 4, we will explore alternative coinsurance schemes that differ in terms of premium and thus in terms of disposable income. The implied differences in disposable income are too small to exert important effects upon risk aversion, however.

To maintain symmetry, we specify a utility function that is quadratic in both non-medical and medical consumption. To assure that the marginal utility of non-medical consumption is positive everywhere, we constrain the range of the parameter β. We use u for the consumer's direct utility, z for the consumption of health care, c for the consumption of non-medical services and y for the consumer's gross income. So we have:

$$\begin{aligned} &u = c - \tfrac{1}{2}\beta c^2 + \gamma z - \tfrac{1}{2}\delta z^2 \\ &0 < \beta < 1/y, \quad \delta > 0 \end{aligned} \tag{1}$$

The parameter γ differs between different states of health. This reflects patient heterogeneity in terms of health status. We interpret this as reflecting a state-dependent health production function. That the marginal utility of health care consumption, $\gamma - \delta z$, is increasing in γ reflects that we assume that the value of medical care is a decreasing function of the health status of the patient: the worse the health of a patient, the more beneficial will be medical intervention. Preferences in (1) are separable. Finkelstein *et al.* (2013) have shown empirical evidence for a state-dependent form in which the marginal utility of non-medical consumption declines when health deteriorates. We will return to this below.

The rate of coinsurance is denoted as b. This coinsurance rate can take any value between zero and one: $0 \leq b \leq 1$. 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 coinsurance payments m, the budget constraint of the consumer is nonlinear:

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

Here, p denotes the uniform health insurance premium.

The consumer acts rationally. Hence, he maximizes (1), subject to (2), given the value of the parameter γ that reflects his health status. 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 our numerical 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 healthy consumer with zero health care demand. Second, γ can be $\gamma_2 > \gamma_1$, for which $\gamma_2 - bt(1-\beta y) > 0$. This characterizes a patient in need of health care and for whom coinsurance payments increase with health spending. Thirdly, γ can be $\gamma_3 > \gamma_2$, which implies that health care demand is sufficiently large to make the consumer pay the maximum of coinsurance payments m. The probabilities attached to the three cases are π_1, π_2 and π_3.

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

$$
\begin{aligned}
z_1 &= 0 \\
z_2 &= \frac{\gamma_2 - bt\left(1-\beta(y-p)\right)}{\delta + \beta\left(bt\right)^2} \\
z_3 &= \frac{\gamma_3}{\delta}
\end{aligned}
\tag{3}
$$

An important feature is that health care demand is price-elastic only in the second state. Demand in states 1 and 3 is inelastic, albeit for different reasons.

We assume that the health insurance industry is perfectly competitive and features zero administration costs. Alternatively, one could model a national insurer, which would give the same results as long as one would keep the utility function and the budget constraint with zero administration costs. (Adverse selection is a major reason for the two market structures to be different. Adverse selection does not play a role in our analysis as we assume a representative agent.) Hence, health insurance premiums equal health spending minus coinsurance payments:

$$p = \pi_2(1-b)tz_2 + \pi_3(tz_3 - m) = \pi_2(1-b)tz_2 + \pi_3(t\frac{\gamma_3}{\delta} - m) \tag{4}$$

Let us use v_i to denote the maximum of utility, *i.e.* u evaluated in the point of optimal health care consumption z_i $i = 1,2,3$. Expected utility, denoted V, weighs utility in the three states:

$$V = \pi_1 v_1 + \pi_2 v_2 + \pi_3 v_3 \tag{5}$$

We now derive the following proposition:

Proposition 1: Maximization of expected utility (equation (5)) with respect to the coinsurance rate b yields a value that obeys $0 < b < 1$, given that π_3 is sufficiently small.

Proof:
We write the derivative of V with respect to b as follows,

$$\frac{dV}{db} = \frac{\partial V}{\partial b} + \frac{\partial V}{\partial z_2}\frac{dz_2}{db} = \frac{\partial V}{\partial b} + \frac{\partial V}{\partial z_2}\varepsilon\frac{z_2}{b} \tag{6}$$

where ε denotes the price elasticity of health care demand in the second state, which can easily be shown to be strictly negative. Appendix B derives that this derivative of V with respect to b can be formulated as follows,

$$\frac{dV}{db} = -\Phi_1\left[\Phi_2 b + \Phi_3\varepsilon\frac{(1-b)}{b}\right] \tag{7}$$

where $\Phi_1 > 0$, $\Phi_2 > 0$ for π_3 sufficiently small and $\Phi_3 > 0$.

If π_3 is sufficiently small, equation (7) tells us the following. First, the derivative dV / db goes to plus infinity when b approaches zero (recall that $\varepsilon < 0$). This suggests that the optimal coinsurance rate, which features $dV / db = 0$, is strictly positive. Similarly, dV / db is strictly negative for $b = 1$ (Appendix B shows that $\Phi_2 > 0$ if π_3 is sufficiently small). This suggests that the optimal coinsurance rate is strictly below one. Combining the two results, it follows that optimal coinsurance implies partial insurance, *i.e.* $0 < b < 1$.

What is the role of the condition that π_3 is sufficiently large? Asssume that π_3 violates this condition. This implies that $\Phi_2 < 0$. This in turn implies that dV / db is positive for $b = 1$. As a consequence, there is no interior solution for b in the domain $0 < b \leq 1$ and the optimal coinsurance rate equals one.

In order to explain the difference in results, note that an increase in the coinsurance rate has three types of welfare effects. First, it combats the moral hazard distortion, which is welfare-increasing. Second, it increases the financial uncertainty in the second state of the model, in which people copay, but less than the maximum. This is welfare-reducing. Thirdly, an increase in coinsurance payments implies transfers from the first and second state to the third state, in which people pay the maximum of coinsurance payments. As the marginal utility of income is highest in this third state, this third effect is to increase welfare.

If π_3 is sufficiently small, this third effect plays a minor role. This is the classical case: an increase in the coiunsurance rate decreases moral hazard and increases financial uncertainty so that the optimal coinsurance rate ends up between zero and one. However, if π_3 is sufficiently large, the total of the three effects of an increase in the coinsurance rate is positive for all $0 < b \leq 1$ so that it is optimal to raise the rate as much as possible. Then, a deductible is optimal: $b = 1$.

We conclude that our model can help to explain the existence of deductibles in real-world health insurance constracts. We do want to push this argument further, however. In the numerical simulations in this paper in which π_3 is determined endogenously, deductibles are absent: in all simulations, the optimal coinsurance rate is strictly below one. Therefore, we put π_3 equal to zero in the remainder of this section.

Based on the result in equation (7), we state our second proposition:

Proposition 2: More price-elastic health care demand (a more negative value for ε) implies a higher optimal coinsurance rate *b*.

Proof:

In order to find the relation between the optimal coinsurance rate and the elasticity of health care demand, if any, we totally differentiate dV/db with respect to b and ε and derive from this an expression for the derivative $db/d\varepsilon$:

$$\begin{gathered}\frac{\partial(dV/db)}{\partial\varepsilon}+\frac{\partial(dV/db)}{\partial b}\frac{db}{d\varepsilon}=0\\ \rightarrow\frac{db}{d\varepsilon}=-\left[\frac{\partial(dV/db)/\partial\varepsilon}{\partial(dV/db)/\partial b}\right]\end{gathered}\tag{8}$$

Appendix B derives that $db/d\varepsilon<0$. Hence, a higher price elasticity of health care demand (a decrease of the negatively signed price elasticity ε) increases the optimal coinsurance rate b.

The result that more price-elastic demand aggravates the moral hazard distortion and thus increases the coinsurance rate that balances moral hazard and financial uncertainty is well known. What proposition 2 adds is that this result also holds true in a model that accounts explicitly for a coinsurance maximum.

The increase of the optimal coinsurance rate upon the introduction of a cap on spending hinges on two elements. The first is that the (absolute value of the) price elasticity of health care demand relates negatively to health care spending. This empirical observation is implied by the quadratic structure of our model (see Appendix A). As a consequence, the introduction of a cap which reduces health care spending for which copayments are required, makes health care demand more price-elastic. The second element is proposition 2: the optimal coinsurance rate increases upon an increase in the (absolute value of the) price elasticity of health care demand. Combining the two elements, it follows that the introduction of a cap on spending increases the optimal coinsurance rate.

To see the distributional effects of the introduction of a copayment maximum and the accompanying increase in the coinsurance rate, let us take a look at Figure 1. This figure displays coinsurance payments as a function of the health status parameter γ. Here, the line 0AB represents coinsurance payments in the linear case. The line 0DAC represents coinsurance payments in the case with a cap: it puts a cap of C on coinsurance payments and features a higher coinsurance rate (the line 0DE is steeper than the line 0AB). The line 0FGH measures the difference between 0DAC and 0AB. It shows that the distribution of coinsurance payments is inverse V-shaped, that the consumer with spending at a level at which the cap starts to apply, is least well-off and that consumers with marginally higher spending are also worse off as they face an increase in coinsurance payments up to the maximum.

INSERT FIGURE 1 ABOUT HERE

3 The simulation model

The previous section discussed a model that was kept stylized in order to be able to explore the properties of the optimal coinsurance rate. This model suffers from two drawbacks. First, it includes three states of nature only, which is not that realistic. Second and related to the first, the model implicitly assumes that patients take the price of medical consumption as a given. In reality however, if the insurance scheme requires coinsurance up to a certain point of health care consumption, patients can reduce the price of their consumption to zero by consuming beyond that point. This section modifies the model of the previous section on both points. As we will see, our main results continue to hold true.

3.1 Consumers

In order to obtain a better match with reality, the more general model in this section also differentiates health care into two groups, outpatient care (O) and all care (A) which comprises of outpatient care and inpatient care. In this more general version, the patient receives two shocks. The first shock indicates what kind of services the patient needs to improve his or her health status. This is outpatient care, all care or zero care (N). The associated probabilities are $(1-\pi_N)\pi_O$, $(1-\pi_N)\pi_A = (1-\pi_N)(1-\pi_O)$ and π_N . The second shock indicates, as before, the size of the intervention that is required to restore the patient's health status. The second shock associates each patient with a particular value for γ . In both the cases of outpatient care and all care, the parameter γ is drawn from a lognormal distribution; in the case of zero care, γ equals zero.

In the cases of outpatient care (O) and all care (A), the consumer faces an optimization problem like in the previous section, *i.e.* decide how much health care to demand given the value of γ . Unlike the model in the previous section, each patient now effectively faces two prices for health services, however.

Our method to solve the optimization problem resembles Hausman's (1985) *maximum maximorum* principle. In a first step, we maximize utility (equation (1)) subject to one of the linear segments of the budget constraint. We repeat this procedure three times, for the two parts of the budget constraint (equation (2)) to find out which of the two interior solutions applies, and for $z = 0$ to find out where the corner solution applies. This results in a health care demand function and an indirect utility function for each of the three linear segments of the budget constraint. In a second step, we compare the three indirect utility functions to find out which function applies to which range of values for γ .

Following this optimization procedure, we derive the following expression for health care demand:

$$
\begin{array}{ll}
z = 0 & 0 \le \gamma \le \gamma_0 \\
z = \dfrac{\gamma - bt\left(1-\beta(y-p)\right)}{\delta + \beta\left(bt\right)^2} & \gamma_0 \le \gamma \le \gamma_1 \\
z = \dfrac{\gamma}{\delta} & \gamma_1 \le \gamma
\end{array}
\tag{9}
$$

The corresponding expressions for the boundary values γ_0 and γ_1 are as follows:

$$
\gamma_0 = bt\left(1-\beta(y-p)\right)
$$

$$
\gamma_1 = \frac{-\delta\left(1-\beta(y-p)\right)}{\beta\left(bt\right)} + \frac{\delta}{\beta\left(bt\right)^2}\sqrt{\Omega}
\tag{10}
$$

where

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

Equation (9) implicitly assumes that $\gamma_1 > \gamma_0$. This can be shown to apply if the coinsurance maximum m exceeds some minimum value. We do not report the counterpart of equation (9) that applies if this condition is not met. This is not relevant: in all the simulations we have made, the coinsurance maximum exceeds the minimum value.

Equation (9) demonstrates that health care demand is a piecewise linear function of γ. The economics behind this equation can be illustrated by following what would happen if we hypothetically increased γ from zero to infinity. A state of perfect health ($\gamma = 0$) obviously implies zero demand. Minor illnesses, defined by $0 < \gamma \le \gamma_0$, produce zero demand too. For these illnesses, the health benefits of medical consumption do not balance the utility value of the price the consumer should pay out of pocket so that the corner solution of zero medical spending applies. This changes for higher values of γ. If $\gamma \ge \gamma_0$, health care demand is positive and increasing in the severity of illness. The relationship is linear, until γ hits γ_1. At $\gamma = \gamma_1$, the consumer is in fact indifferent between consuming a small amount of medical care at the marginal out-of-pocket price bt and a larger amount at a zero marginal price. The drop in the out-of-pocket price of health services that occurs when health care expenditure hits the ceiling m necessitates an equally-sized drop in the marginal benefit of medical consumption. This makes health care demand jump at $\gamma = \gamma_1$. For $\gamma > \gamma_1$, demand is a continuous increasing function of γ.

The expression for indirect utility is structured in exactly the same way as that for health care demand: three equations, marked by the same critical values for γ:

$$
\begin{aligned}
v &= (y-p) - 0.5\beta(y-p)^2 && 0 \le \gamma \le \gamma_0 \\
v &= \frac{\delta(y-p) + 0.5(bt)^2 - 0.5\beta\delta(y-p)^2 + 0.5\gamma^2 - \gamma bt(1-\beta(y-p))}{\delta + \beta(bt)^2} && \gamma_0 \le \gamma \le \gamma_1 \qquad (11) \\
v &= \frac{\delta(y-p-m) - 0.5\beta\delta(y-p-m)^2 + 0.5\gamma^2}{\delta} && \gamma_1 \le \gamma
\end{aligned}
$$

Having laid out the general principle of optimization that applies both to consumers of all (A) care and to consumers of outpatient (O) care, we now point to a difference. We denote the minimum value for γ, the severity of the health shock, as $\gamma_{\min}$. This $\gamma_{\min}$ value differs between the group of A care and the group of O care. In the case of O care, $\gamma_{\min}$ equals zero. Equations (9) to (11) hold unambiguously. In the case of A care, $\gamma_{\min} > 0$ (as A care includes at least one hospital visit, $\gamma_{\min} > 0$ allows for a better calibration of the model). Now, three cases are possible: (i) $\gamma_{\min} < \gamma_0$. Equations (9) to (11) apply, except that the first segment is $\gamma_{\min} \le \gamma \le \gamma_0$, rather than $0 \le \gamma \le \gamma_0$. (ii) $\gamma_0 < \gamma_{\min} < \gamma_1$. The first segment of equations (9) and (11) does not apply and the second segment is $\gamma_{\min} \le \gamma \le \gamma_1$, rather than $\gamma_0 \le \gamma \le \gamma_1$. (iii) $\gamma_{\min} > \gamma_1$. Only the third segment of equations (9) and (11) applies. This case does not occur in any of our simulations.

Our analysis distinguishes bounded from unbounded coinsurance schemes. Bounded schemes feature a finite coinsurance maximum. In this case, equations (9) to (11) apply.

Unbounded schemes do not put a maximum on coinsurance payments, or, equivalently, impose a maximum that is infinitely high. It can be derived from equation (10) that, in this case, γ_1 is infinite and, thus, the third segments of equations (9) and (11) do not apply.

3.2 Health insurers

For both the O and A groups of health care, we integrate the levels of consumer utility for all values of γ to arrive at expressions for expected utility. Letting *G*(.) denote the distribution function of γ_i (i = O, A), we obtain the following:

$$E(v_i) = G(\gamma_{i,0})\, E(v_i \mid \gamma_i \leq \gamma_{i,0}) + \big(G(\gamma_{i,1}) - G(\gamma_{i,0}) \big) E(v_i \mid \gamma_{i,0} \leq \gamma_i \leq \gamma_{i,1}) + \\ \big(1 - G(\gamma_{i,1})\big) E(v_i \mid \gamma_i \geq \gamma_{i,1}) \quad i = O, A \tag{12}$$

Expressions for the conditional expectation variables $E(v \mid .)$ in equation (12) are based upon the expressions in equation (11). Expressions for both $G(.)$ and $E(v \mid .)$ in equation (12) reflect the lognormality of the distributions for γ and can be found in Appendix C.

Aggregate expected utility can now be written as follows,

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

where v_N, utility in case of zero care, reads as $(y-p) - \frac{1}{2}\beta(y-p)^2$. Equation (13) then calculates aggregate expected utility as the weighted average of expected utility in case of zero care (*N*), outpatient care (*O*) and all care (*A*).

In order to complete our model, we need to develop an expression for the health insurance premium p. This premium equals the difference between the aggregates of spending and coinsurance payments, which we will denote as k and w respectively. Given a population mass of unity, these variables equal their expected values, $E(k)$ and $E(w)$.

$$p = E(k) - E(w) \tag{14}$$

Expected health care spending can be elaborated as follows,

$$E(k) = (1-\pi_N)\pi_O E(k_O) + (1-\pi_N)\pi_A E(k_A) \tag{15}$$

with

$$\begin{aligned} E(k_i) = &\left(G\left(\gamma_{i,1} \right) - G\left(\gamma_{i,0} \right) \right) t_i \left[E\left(z_i \mid \gamma_{i,0} \le \gamma_i \le \gamma_{i,1} \right) \right] \\ &+ \left(1 - G\left(\gamma_{i,1} \right) \right) t_i \left[E\left(z_i \mid \gamma_i \ge \gamma_{i,1} \right) \right] \qquad i = O, A \end{aligned} \tag{16}$$

Aggregate coinsurance payments are defined in a similar way,

$$E(w) = (1-\pi_N)\pi_O E(w_O) + (1-\pi_N)\pi_A E(w_A) \tag{17}$$

with

$$\begin{aligned} E(w_i) = &\left(G\left(\gamma_{i,1} \right) - G\left(\gamma_{i,0} \right) \right) b t_i \left[E\left(z_i \mid \gamma_{i,0} \le \gamma_i \le \gamma_{i,1} \right) \right] \\ &+ \left(1 - G\left(\gamma_{i,1} \right) \right) m \qquad i = O, A \end{aligned} \tag{18}$$

The conditional expectation variables $E\left(z_i \mid .\right)$ in equations (16) and (18) can be derived from the expressions in equation (3). See Appendix C for details.

We are now almost ready to turn to the model simulations. In these simulations, the health insurer maximizes aggregate expected utility as defined in equation (13), using the coinsurance rate b and, in case of a bounded insurance scheme, the maximum m as instruments. However, we first need to fill the model, *i.e.* find values for the parameters of our model.

4 Numerical simulations

4.1 Calibration of the model

We use calibration to obtain values for the parameters of our model. In particular, we calibrate on data for the privately insured population in the Netherlands (before the 2006 health insurance reform). We define inpatient services as services related to hospitalization and outpatient services as services provided by the general practitioner, the physiotherapist, pharmaceuticals and outpatient specialist services. As indicated above, for the two groups of medical services O and A, $\ln(\gamma - \gamma_{\min})$ is assumed to be normally distributed; the value $\gamma_{\min}$ is zero in case of the O group and positive in case of the A group.

In total, we need to find values for 12 parameters. Two of these are probabilities: the probability of zero need for medical services, π_N, and the conditional probability of need for both inpatient and outpatient services, π_A. The latter is determined by using data on the frequency of a hospital admission. Two parameters are insurance policy variables: b and m. As regards b, we put it equal to one, reflecting the deductible that features the insurance of our sample population. Five parameters describe the moments of the two lognormal distribution functions for the γ_O and γ_A parameter ($\mu_O, \mu_A, \sigma_O, \sigma_A$) and the minimum value $\gamma_{\min}$ for the A group. We base values for σ_O and σ_A on the coefficients of variation of the distributions of health care spending in the O and A case respectively and a value for $\gamma_{\min}$ on an estimate of the minimum of health spending in case of a hospital admission.

This leaves four parameters to be determined. In addition, we have to find values for the three parameters that define the utility function: β, δ_O and δ_A. Values for these seven parameters are computed simultaneously. Particularly, they are calculated such that the predicted values of the probability of zero health care spending, of aggregate coinsurance payments and of health spending on O services and A services match the data, that the predicted coefficient of relative risk aversion (evaluated at the mean of non-medical consumption) equals

2 (Garber and Phelps 1997), and that the price elasticities of demand for O services and A services (evaluated at the mean of health consumption) match the estimates reported in Van Vliet (2001).

The features of the calibrated model are quite standard. The probability of spending on outpatient care is about ten times as high as the probability of spending on all care. However, the distribution function for health spending of all care features a higher minimum spending, a higher mean and is more skewed to the right than the distribution function for health spending on outpatient care. Hence, the spending on all care services is almost double that on outpatient care services. The price of a unit of all care is also higher than the price of a unit of outpatient care. Hence, aggregate spending on all care services is more than 16 times as high as aggregate spending on outpatient care services.

The price elasticity of medical services (evaluated at average health spending) equals 0.079 for outpatient care and 0.011 for the combination of inpatient and outpatient care (in absolute terms). Although based on Van Vliet (2001) and in line with a large empirical literature, they are at the lower end of the price elasticities as obtained in the RAND Health Insurance Experiment (Newhouse *et al.* 1993). Below, we will explore the sensitivity of our results with respect to these price elasticities.

4.2 The optimal unbounded coinsurance scheme

Having calibrated the model, we now turn to simulations. We start with simulating an unbounded scheme, *i.e.* a scheme without a cap on coinsurance payments.

Which value of the rate of coinsurance maximizes expected household utility in the absence of a cap? The model is too complex to explore this question analytically. We therefore apply a grid search procedure in which we let the coinsurance rate run from 0% to 100% in steps of 5 percentage points. The model requires that average patient income, net of health care premiums and co-payments is positive for every value of γ. Therefore we put an upper bound on coinsurance payments at a level of 50,000 euro. Given that the probability mass of health spending exceeding 50,000 euro is only 0.02%, the approximation error involved is negligible.

Indeed, calculations with a maximum of coinsurance payments of 40,000 euro (not shown) yield very similar outcomes.

Table 1 displays our results. The first number in its first column shows that the optimal coinsurance rate is 30%. From the inspection of all grid points, the optimum is globally concave, a result that carries over to all other simulations. As expected, this value is in between zero and one. The optimal policy strikes a balance between the benefits from insurance, *i.e.* smoothing of non-medical and medical consumption across health states, and the cost of insurance, *i.e.* a high insurance premium due to the moral hazard in medical consumption.

In order to see how much this numerical result depends upon the assumed parameter values, Table 1 also shows the results for cases that feature much smaller or larger values for 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 are intuitive. A low risk aversion (a low value for β), high price elasticities of demand (low values for δ_O and δ_A), low standard deviations of the health status (low values for σ_O and σ_A) and a high share of outpatient care in health care (a high value for π_O) correspond with a high coinsurance rate and vice versa. Quantitatively, the optimal coinsurance rate displays quite some variation: it varies between 10% and 55%. Broadly speaking, these results are in line with Manning and Marquis (1996) who calculate the optimal coinsurance rate to lie in the range from 40 to 50%.

INSERT TABLE 1 ABOUT HERE

One way to see what optimality means is to calculate the income loss that a person under an optimal scheme would be willing to accept in order to prevent being moved to a different scheme, *e.g.* a scheme without any insurance or a scheme with full insurance. The second and third columns of Table 1 display the compensating variations of zero insurance and full insurance. We define the compensating variation of an alternative scheme (zero insurance or full insurance or any other scheme) as the income that a household under the optimal scheme would be willing to pay in order to leave him indifferent between the optimal scheme and the alternative scheme. Formally, the compensating variation of alternative scheme Q, $\tilde{y}^Q$, is defined by the equality $V(\hat{b}, \hat{m}, y - \tilde{y}^Q) = V(b^Q, m^Q, y)$, where $\hat{b}$ and $\hat{m}$ denote the coinsurance rate and maximum of the optimal scheme and b^Q and m^Q denote the counterparts of alternative scheme Q. We use $\tilde{y}^{ZI}$ and $\tilde{y}^{FI}$ to denote the compensating variations of the schemes of zero insurance and full insurance respectively.

The compensating variation of the welfare loss from zero insurance is 99 euro, that from full insurance is smaller, namely 21 euro. These results are interesting on two accounts. First, given that we compare the optimal scheme with two polar schemes, these numbers are pretty small. Second, according to this calculation, full insurance is superior to no insurance. This may seem surprising, as Feldstein (1973), Feldman and Dowd (1991) and Manning and Marquis (1996) find the opposite result. The result is not robust to parameter changes, though. Indeed, the simulations in Table 1 show that both full insurance and zero insurance may dominate the other ona and that which case arises depends upon the assumed degree of risk aversion and the assumed price elasticities of health care demand.

4.3 The optimal bounded coinsurance scheme

What are the implications of introducing a maximum to coinsurance payments and choosing optimally both the coinsurance rate and this maximum ? The instrument of a cap on coinsurance payments helps to reduce coinsurance payments for those in very bad states of health. If the insurer, as before, also chooses the optimal rate of coinsurance, the introduction of a cap to coinsurance payments allows the insurer to fine tune the coinsurance rate to the price elasticity of health care demand in the states with better health.

In line with the approach adopted for the case of the optimal unbounded scheme, we explore which combination of coinsurance rate and maximum achieves the highest level of expected utility. Our grid search procedure varies the coinsurance rate from 0% to 100% in steps of 5 percentage points and the maximum of coinsurance payments from a minimum of 100 euro to a maximum of 50,000 euro in steps of 50 euro. The coinsurance maxima that characterize the optimal schemes in our simulations deviate strongly from both these minimum and maximum values.

Table 2 summarizes our results. The optimal coinsurance maximum is about 3,250 euro, the optimal coinsurance rate equals 55%. The health care reform that introduces a cap on coinsurance payments thus increases the coinsurance rate from 30 to 55%. It increases coinsurance payments by about 40%. Aggregate health expenditure and health insurance premiums fall. Also interesting is what happens to coinsurance payments in the groups of outpatient and all care consumers. Not surprisingly, coinsurance payments in the states of outpatient care increase. What is surprising, at least at first sight, is that coinsurance payments by all care consumers increase as well. Although the group of all care consumers have high spending due to hospitalization, a sizeable fraction of the consumers of all care services has spending below the level at which the maximum starts to apply. These people face an increase of coinsurance payments on account of a higher coinsurance rate. This effect dominates the direct effect of a cap on coinsurance payments.

The optimal bounded scheme features a declining coinsurance rate: the degree of insurance offered by the optimal scheme is higher for those with high health spending than for those with low health spending. The same result can be found in earlier work, particularly Arrow (1963), Drèze and Schokkaert (2013) and Blomqvist (1997). Like Blomqvist (1997) but unlike Arrow (1963) and Drèze and Schokkaert (2013), the optimal coinsurance rate is below one, *i.e.* the optimal scheme features insurance also for health spending below the cap.

INSERT TABLE 2 ABOUT HERE

The compensating variation of the welfare gain that corresponds to capping coinsurance is very modest: 9 euro per insured only. This is in line with the results found earlier for the case of full insurance and zero insurance. These two cases feature welfare measures of 99 euro and 21 euro respectively, and moving from the optimal capped scheme to one of these schemes would imply a bigger reform than moving from the optimal capped scheme to the optimal linear scheme.

Moreover, the aggregate welfare gain hides substantial transfers between different groups of consumers. Those who consume all care services loose 201 euro, whereas the consumers of outpatient services gain 29 euro. The former group suffers relatively much from the increase in the coinsurance rate, wheras the two groups share the gain from lower health insurance premiums.

Let us look more closely to the group of all care consumers. The loss does not occur to all of them. The population density below γ_1 is 84 percent. Hence, the cap on coinsurance payments benefits only 16 percent of the group, Happily, these 16 percent are the most needy health care consumers. The majority consumers in group A loose, however. The patient with health care spending at the level at which the coinsurance maximum starts to apply ($z = m / (bt)$, hence $\gamma = \gamma_1$) bears the maximum loss. This amounts to 1,317 euro.

Table 3 displays the results of a sensitivity analysis on the introduction of a cap on coinsurance. The optimal coinsurance rate ranges between 45% and 90%, the optimal coinsurance maximum between 1,050 and 7,500 euro and the corresponding compensating variation between 0 and 22 euro. These results are in line with the benchmark simulation: the introduction of a maximum to coinsurance payments entails an increase of the coinsurance rate and a modest aggregate welfare gain. The effect on the consumers of all care services varies widely. In 4 out of 9 cases (see Table 3), the consumers of all care services benefit and in another 4 cases, they loose from the reform.

The welfare effects for the groups of outpatient care consumers and all care consumers depend upon the maximum that the capped insurance scheme imposes and the increase in the coinsurance rate that correspond to a typical simulation. To take just one example, suppose β is 1.5 times as large as in the benchmark simulation. This corresponds to a relatively high degree of risk aversion. Quantitatively, the results differ strongly from those in the benchmark case. Now, the consumers of all care benefit (641 euro in compensating variation terms) and the consumers of outpatient care services loose (48 euro in compensating variation terms). The optimal scheme now involves a maximum to coinsurance payments of 1,100 euro rather than 3,250 euro. This implies a bigger welfare gain for those with high health care spending.

Furthermore, the optimal scheme features an increase in the coinsurance rate of 65 percentage points rather than 25 percentage points as in the benchmark case. Consumers of outpatient care services are worse off, on account of the steep increase of the coinsurance rate and due to the fact that a maximum of 1,100 euro requires a large increase in premiums.

INSERT TABLE 3 ABOUT HERE

4.4 Decomposing the effects of the policy reform

The policy reform that puts an upper limit to coinsurance payments can analytically be split into two component reforms. One of them puts a cap on coinsurance payments without changing the coinsurance rate. The second one keeps the coinsurance maximum at its new level and increases the coinsurance rate. An analysis of the two component reforms allows us to see more clearly the different effects that determine the overall effect of the insurance reform.

Putting a cap on coinsurance payments increases aggregate welfare with 5 euro (in terms of Table 4, the difference between 9 and 4 euro). This is about half the welfare gain of the reform. From the introduction of such a hybrid scheme, the consumers of all services would gain 348 euro and the consumers of outpatient services would lose 28 euro. This confirms the intuition that the introduction of a cap on coinsurance payments without raising the coinsurance rate benefits consumers with high levels of spending who are overrepresented in the A group. The increase in health insurance premiums which has to make up for the reduction of coinsurance payments has to be borne by all, which explains the welfare loss suffered by outpatient services consumers.

The second part of the reform then increases the coinsurance rate from 30% to 55%. This benefits the average person 4 euro. It makes the average outpatient care services consumer gain 57 euro and the average all care services consumer loose -549 euro. The consumers of inpatient services loose as they suffer disproportionately from the increase in the coinsurance rate, but share with the outpatient care services consumers the reduction of the health insurance premium.

INSERT TABLE 4 AND 5 ABOUT HERE

4.5 The least well-off consumers

Above, we calculated the introduction of a cap in the benchmark parameter case to imply a welfare gain of 9 euro at the macro level. At the level of outpatient care consumers, a gain of 29 euro was found, whereas the consumers of all care services were put at a loss of 201 euro. Among these two groups, there is a lot of heterogeneity however. In both groups, the patient that has spending at the level at which the coinsurance payments achieve their maximum value, is hurt most. The loss for this least well-off outpatient care services consumer amounts to 1,343 euro, wheras the loss for the least well-off all care services consumer amounts to 1,317 euro.

That the losses for the least well-off consumers are so big relates to the increase of the coinsurance rate that accompanies the introduction of a cap. To show this more clearly, we have calculated the maximum loss in case of the hybrid scheme discussed in the previous section, *i.e* a capped scheme that has the original coinsurance rate. The maximum loss is now 30 euro only, which contrasts sharply with the amount of 1,343 euro.

These large numbers for the least well-off consumers are confirmed in our sensitivity analysis. Table 5 reports these numbers for the alternative simulations included in Table 6. Although the maximum loss may be negligible (witness the $\sigma_i = 0.50\sigma_{i,BM}$ simulation), most of the numbers are similar to that of the benchmark simulation.

5 Concluding remarks

Overall, the numerical simulations in this paper exhibit quite some variation. Given the uncertainty about the true values of the parameters, we cannot come up with precise estimates of the numerical effects of a cap on coinsurance payments. On a qualitative level, our results are robust to parameter changes, however. In particular, our result that the introduction of a cap on coinsurance payments has a relatively small efficiency effect, that it increases the coinsurance rate and that it (thus) implies relative large losses for some individuals, especially those with spending close to the level at which coinsurance payments reach their maximum level, are found in all alternative simulations investigated.

Obviously, our results may be challenged by making different assumptions. In particular, one could argue for a different kind of utility function (like the state-dependent utility function in Finkelstein *et al.* (2008), *e.g.*), a role for the suppliers of medical services, or some other assumption on the information the patient has available when visiting the doctor. Our qualitative results will very likely survive, however, as long as one assumes that the price elasticity of health care demand declines (in absolute terms) when health status deteriorates and that the coinsurance rate as chosen by the insurer, is positively related to the price elasticity of health care demand (in absolute terms) as suggested by a long-standing literature in health economics.

The paper could be extended in several ways. It would be interesting to further explore the case for deductibles. The paper already hinted at the idea that deductibles may be optimal policies. In addition, it would be interesting to release the assumption that health insurers are completely uninformed and allow them to use this information to differentiate the parameters of the insurance scheme with respect to different types of services. Particularly, if demand for some services is more price-elastic than demand for other services and different types of services cannot be seen as substitutes, differentiation of the parameters of the insurance scheme can be used to increase welfare further.

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.

References

ABRAMOWITZ, M. and I.A. STEGUN, 1972. (eds), Handbook of Mathematical Functions, Dover Publications, Mineola N.Y.

ARROW, K.J., 1963. Uncertainty and the Welfare Economics of Medical Care, *American Economic Review* **53**, 941-973.

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**, 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.

CHANDRA, A. and J. SKINNER, 2012. Technology Growth and Expenditure Growth in Health Care, *Journal of Economic Literature* **50**, 645-680.

CHANDRA, A., J. HOLMES and J. SKINNER, 2013. Is This Time Different? The Slowdown in Health Care Spending, *Brookings Papers on Economic Activity*, 261-323.

CUTLER, D.M. and R.J. ZECKHAUSER, 2000. The Anatomy of Health Insurance, in: A.J. Culyer and J.P. Newhouse (eds.), *Handbook of Health Economics*, vol. 1A, Elsevier, 563-643.

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.

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.

ELLIS, R.P. and W.G. MANNING, 2007. Optimal Health Insurance for Prevention and Treatment, *Journal of Health Economics* **26**, 1128-50.

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.

FEENBERG, D. and J. SKINNER, 1994. The Risk and Duration of Catastrophic Health Care Expenditure, *Review of Economics and Statistics* **76**, 633-647.

FELDMAN, R. and B. DOWD, 1991. A New Estimate of the Welfare Loss of Excess Health Insurance, *American Economic Review* **81**, 297-301.

FELDSTEIN, M.S., 1973. The Welfare Loss of Excess Health Insurance, *Journal of Political Economy* **81**, 251-80.

FELDSTEIN, M. and J. GRUBER, 1995. A Major Risk Approach to Health Insurance Reform, in J. Poterba (ed.), Tax Policy and the Economy 9, NBER, MIT Press, 103-130.

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.

HALL, R.E. and C.I. JONES, 2007. The Value of Life and the Rise in Health Spending, *Quarterly Journal of Economics* **122**, 39-72.

HAUSMAN, J.A., 1985. The Econometrics of Nonlinear Budget Sets, *Econometrica* **53**, 1255 - 82.

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.

McGUIRE, T.G., 2012. Demand for Health Insurance, in M.V. PAULY, T.G. McGUIRE and P.P. BARROS (eds.), Handbook of Health Economics, Elsevier, vol. 2, 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.

OECD, 2011. Health at a Glance 2011, OECD indicators, doi: 10.1787/19991312.

OECD, 2015. Health policies and data, Cost-sharing requirements and exemptions of copayments 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.

STATISTICS NETHERLANDS, Statline online data base, 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., 2014. Population Ageing and Health Care Expenditure Growth, in S. Harper and K. Hamblin (eds.), *International Handbook on Ageing and Public Policy*, Edward Elgar, 178-190.

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.

Figure 1: Coinsurance payments in the linear scheme and the scheme with a cap

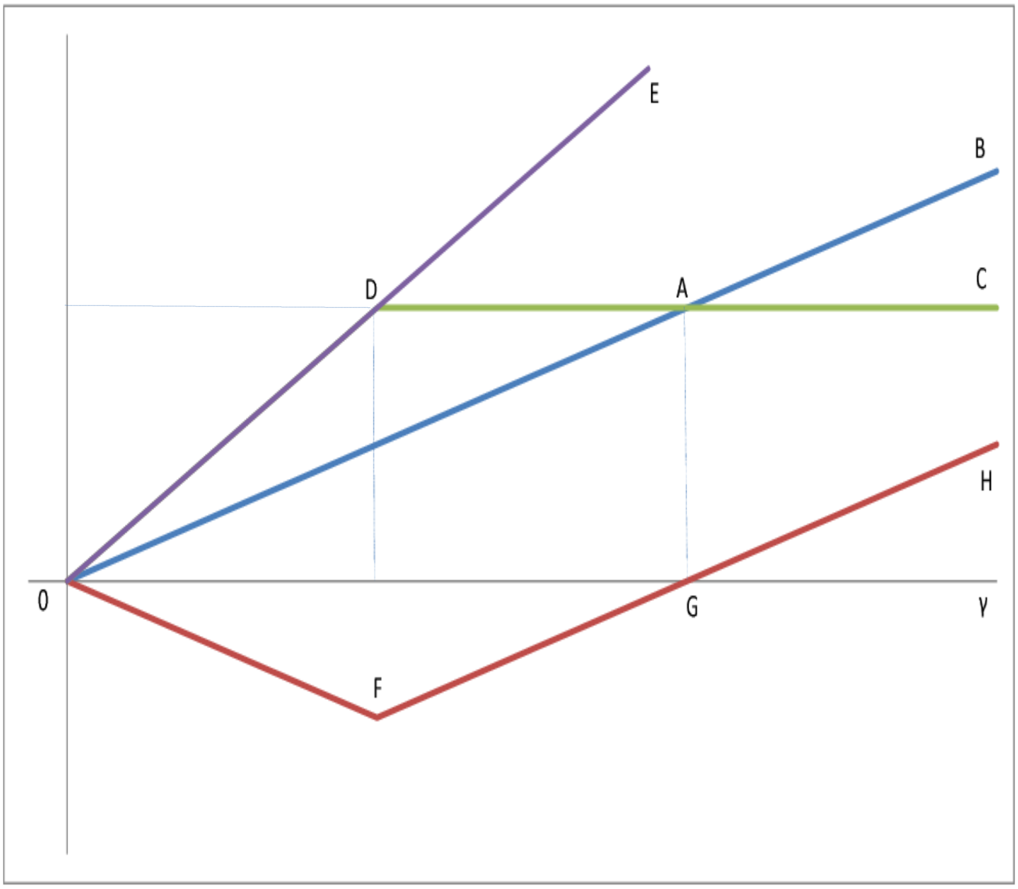

Table 1 Sensitivity analysis optimal unbounded coinsurance scheme

	$\hat{b}$(%)	$\tilde{y}^{ZI}$	$\tilde{y}^{FI}$
Benchmark (BM)	30	99	21
$\beta = 0.50\beta_{BM}$	55	30	58
$\beta = 1.50\beta_{BM}$	20	178	9
$\delta_i = 0.50\delta_{i,BM}$	45	54	64
$\delta_i = 1.50\delta_{i,BM}$	20	115	10
$\sigma_i = 0.50\sigma_{i,BM}$	45	43	32
$\sigma_i = 1.50\sigma_{i,BM}$	10	201	8
$\pi_O = 0.95\pi_{O,BM}$	30	148	24
$\pi_O = 1.05\pi_{O,BM}$	40	46	19

$i = O, A$

Table 2 Effects of reforming the coinsurance scheme

	A	B	$(B-A)/A$ (%)
Optimal coinsurance rate (%)	30	55	
Optimal coinsurance maximum	50,000	3,250	
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
Coinsurance payments	282.67	395.50	39.92
Coinsurance payments, O group	132.55	224.42	69.31
Coinsurance payments, A group	2,009.03	2,362.99	17.62
Insurance premium	659.60	529.57	-19.71

A: The optimal unbounded coinsurance scheme

B: The optimal bounded coinsurance scheme

Table 3 Sensitivity analysis on optimal bounded coinsurance scheme

	$\hat{b}(\%)$		$\hat{m}$	$\tilde{y}^A$	$\tilde{y}_O^A$	$\tilde{y}_A^A$
	A	*B*				
Benchmark (BM)	30	55	3,250	9	29	-201
$\beta = 0.50\beta_{BM}$	55	70	7,100	8	8	0
$\beta = 1.50\beta_{BM}$	20	85	1,100	11	-48	641
$\delta_i = 0.50\delta_{i,BM}$	45	65	5,500	12	22	-87
$\delta_i = 1.50\delta_{i,BM}$	20	85	1,150	8	-44	582
$\sigma_i = 0.50\sigma_{i,BM}$	45	45	7,500	0	-1	8
$\sigma_i = 1.50\sigma_{i,BM}$	10	60	2,450	22	95	-787
$\pi_O = 0.95\pi_{O,BM}$	30	45	3,500	10	27	-98
$\pi_O = 1.05\pi_{O,BM}$	40	90	1,100	10	-65	1961

$i = O, A$

A: The optimal unbounded coinsurance scheme

B: The optimal bounded coinsurance scheme

Table 4 Decomposing the reform of the insurance scheme

	$\hat{b}(\%)$	$\hat{m}$	$\tilde{y}^A$	$\tilde{y}_O^A$	$\tilde{y}_A^A$
Optimal unbounded scheme	30	50,000	9	29	-201
Hybrid scheme	30	3,150	4	58	-565
Optimal bounded scheme	55	3,250			

Table 5 Sensitivity analysis of loss of the least well-off patients

	$\hat{b}(\%)$		$\hat{m}$	$\tilde{y}^A$	$\max(\tilde{y}_O^A)$	$\max(\tilde{y}_A^A)$
	A	*B*				
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

$i = O, A$

A: The optimal unbounded coinsurance scheme

B: The optimal bounded coinsurance 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 (12)):

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

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 $\varepsilon \equiv \left(\partial z/\partial(bt)\right)/(z/(bt))$:

$$\varepsilon = \frac{-2\gamma\beta(bt)^2 - \left(1-\beta(y-p)\right)bt(\delta-\beta(bt)^2)}{\gamma-\left(1-\beta(y-p)\right)bt} \tag{A1}$$

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

$$\varepsilon = -2\beta(bt)^2 - \frac{\left(1-\beta(y-p)\right)bt(\delta+\beta(bt)^2)}{\gamma-\left(1-\beta(y-p)\right)bt} \tag{A2}$$

From this expression, it can easily be derived that $\partial\varepsilon/\partial\gamma > 0$. 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 with a minus sign. Hence, upon an increase in bt, ε becomes more negative, *i.e.* $\partial\varepsilon/\partial(bt) < 0$.

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:

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

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

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

$$\frac{\partial\varepsilon}{\partial(tz)} = \frac{\partial\varepsilon}{\partial\gamma} \Bigg/ \frac{\partial(tz)}{\partial\gamma}$$

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

Appendix B: Derivation of propositions 1 and 2

Recall equation (5) from the main text which defines expected utility:

$$V = \pi_1 v_1 + \pi_2 v_2 + \pi_3 v_3 \tag{5}$$

Using equations (1), (2) and the expressions for z_1 and z_3 yields an expression in terms of p and z_2:

$$\begin{aligned} V = \pi_1[(y-p) - \frac{1}{2}(y-p)^2] + \\ \pi_2[(y-p-btz_2) - \frac{1}{2}\beta(y-p-btz_2)^2 + \gamma_2 z_2 - \frac{1}{2}\delta z_2{}^2] + \\ \pi_3[(y-p-m) - \frac{1}{2}\beta(y-p-m)^2 + \frac{\gamma_3^2}{2\delta}] \end{aligned} \tag{B1}$$

In order to derive an expression for expected utility in terms of z_2 only, we have to eliminate p. Using equation (4), we derive that $y - p = \hat{y} - \pi_2(1-b)tz_2$, where $\hat{y}$ is a shorthand notation for $y - \pi_3(t\gamma_3 / \delta - m)$. Substitution into (B1) yields the following expression:

$$\begin{aligned} V = (\hat{y} - \frac{1}{2}\beta\hat{y}^2) - \pi_2 t(1-\beta\hat{y})z_2 + \pi_2\gamma_2 z_2 - \frac{1}{2}\pi_2\delta z_2^2 + \\ \frac{1}{2}\beta\pi_2(1-\pi_2)(1-b^2)t^2 z_2^2 - \frac{1}{2}\pi_2\beta t^2 z_2^2 - \pi_3 m - \\ \frac{1}{2}\pi_3\beta m^2 + \pi_3\beta\hat{y}m - \pi_3\pi_2\beta(1-b)tz_2 m + \pi_3\frac{\gamma_3^2}{2\delta} \end{aligned} \tag{B2}$$

Differentiation of expected utility in equation (B2) with respect to b yields the following result:

$$\frac{\partial V}{\partial b} = -\pi_2(1-\pi_2)\beta bt^2 z_2^2 + \pi_2\pi_3\beta tz_2 m \tag{B3}$$

Differentiation of expected utility in equation (B2) with respect to z_2 yields the following result:

$$\frac{\partial V}{\partial z_2} = -\pi_2 t(1-\beta\hat{y}) + \pi_2\gamma_2 - \pi_2\delta z_2 - \pi_2\beta t^2 z_2 + \pi_2(1-\pi_2)\beta t^2 z_2(1-b^2) - \pi_2\pi_3\beta(1-b)tm \quad \text{(B4)}$$

The next step is to use equations (B3) and (B4) in order to derive an expression for the derivative dV/db:

$$\begin{aligned}
&\frac{dV}{db} = \frac{\partial V}{\partial b} + \frac{\partial V}{\partial z_2}\frac{dz_2}{db} = \\
&\frac{\partial V}{\partial b} + \frac{\partial V}{\partial z_2}\varepsilon\frac{z_2}{b} = \\
&\left\{-\pi_2(1-\pi_2)\beta b t^2 z_2^2 + \pi_2\pi_3\beta t z_2 m\right\} + \\
&\left\{-\pi_2 t(1-\beta\hat{y}) + \pi_2\gamma_2 - \pi_2\delta z_2 - \pi_2\beta t^2 z_2 + \pi_2(1-\pi_2)\beta t^2 z_2(1-b^2) - \pi_2\pi_3\beta(1-b)tm\right\}\varepsilon\frac{z_2}{b}
\end{aligned} \quad \text{(B5)}$$

Finally, combining equations (3) and (4), we derive a reduced-form expression for z_2,

$$z_2 = \frac{\gamma_2 - bt\left(1-\beta\hat{y}\right)}{\delta + \beta\left(bt\right)^2 + \beta b\left(1-b\right)\pi_2 t^2} \quad \text{(B6)}$$

If we substitute the RHS of equation (B6) for z_2 into equation (B5), we arrive at equation (7) in the main text, which we recall here for convenience.

$$\frac{dV}{db} = -\Phi_1\left[\Phi_2 b + \Phi_3\varepsilon\frac{(1-b)}{b}\right] \quad (7)$$

Corresponding to equation (7) are the following three functions to which the main text refers:

$$\Phi_1 = \frac{\pi_2(\gamma_2 - bt(1-\beta\hat{y}))}{(\delta + \beta(bt)^2 + \beta\pi_2 b(1-b)t^2)^2} > 0$$

$$\Phi_2 = \beta(1-\pi_2)t^2(\gamma_2 - bt(1-\beta\hat{y})) - \beta\pi_3 t(m/b)(\delta + \beta(bt)^2 + \beta\pi_2 b(1-b)t^2)$$

$$\Phi_3 = \pi_2\gamma_2\beta t^2 + t(1-\beta\hat{y})\delta + (1-\pi_2)(1-\beta\hat{y})\beta b^2 t^3 + \beta\pi_3 tm(\delta + \beta(bt)^2 +$$

$$\beta\pi_2 b(1-b)t^2) > 0$$

To derive proposition 2, we recall the expression for $db/d\varepsilon$, which can be derived from totally differentiating equation (7) under the assumption that $\pi_3 = 0$:

N

$$\frac{db}{d\varepsilon} = -\left[\frac{\partial(dV/db)/\partial\varepsilon}{\partial(dV/db)/\partial b}\right] \qquad \text{(B7)}$$

For the numerator and denominator of the term at the RHS of (B7), the following expressions hold true:

$$\partial(dV/db)/\partial\varepsilon = -(1-b)t\{(1-\beta\hat{y})\delta + \beta t\gamma_2\pi_2 + \beta b^2 t^2(1-\pi_2)(1-\beta\hat{y})\} < 0$$

$$\partial(dV/db)/\partial b = \beta bt^2(1-\pi_2)\{-2\gamma_2 + 3bt(1-\beta\hat{y})\} + \varepsilon t(1-\beta\hat{y})\delta$$
$$-\varepsilon\beta t^2\{(2-3b)bt(1-\pi_2)(1-\beta\hat{y}) - \gamma_2\pi_2\}$$

The assumption $\gamma_2 \gg bt(1-\beta y)$ implies that $-2\gamma_2 + 3bt(1-\beta\hat{y}) < 0$ and $(2-3b)bt(1-\pi_2)(1-\beta\hat{y}) - \gamma_2\pi_2 < 0$. Upon substitution, this yields that $\partial(dV/db)/\partial b < 0$. Combining the signs of $\partial(dV/db)/\partial\varepsilon$ and $\partial(dV/db)/\partial b$, it follows that $db/d\varepsilon < 0$.

Appendix C: Exact forms for distribution functions and conditional expectations variables

From equation (14) in section 3.1 and equation (15) in 3.2 it follows that expected utility E(V(..)) depends on conditional moments of the distribution of the health status variable γ. Section 4.1 explains that the distribution of γ is a mix of three distributions, of which two are lognormal distributions and the third is a mass point at value zero. Hence, to compute moments of the distribution of health status requires expressions for conditional expectations of the lognormal distribution function.

A stochastic variable x is lognormally distributed if its logarithm $\ln x$ is normally distributed (with parameters, say, μ and σ). Then the density function $g(.)$ of x obeys:

$$g(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\ln x - \mu)^2}{2\sigma^2} \right) \tag{C1}$$

Now we first derive a general expression for the n-th conditional moment of the distribution of x, $E(x^n \mid a \leq x \leq b)$:

$$E\left(x^n \mid a \leq x \leq b\right) = \int_a^b \frac{x^n}{x\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\ln x - \mu)^2}{2\sigma^2} \right) dx =$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{\ln a}^{\ln b} x^n \exp\left(\frac{-(\ln x - \mu)^2}{2\sigma^2} \right) d\ln x = \tag{C2}$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{\ln a}^{\ln b} \left[\exp(n \ln x)\right] \exp\left[-0.5 \left(\frac{-(\ln x - \mu)^2}{\sigma^2} \right) \right] d\ln x$$

Now the integrand in (C2) can be rewritten as:

$$\exp\left(n\ln x\right)\exp\left(\frac{-(\ln x-\mu)^2}{2\sigma^2}\right)=\exp\left[-\frac{1}{2\sigma^2}\left((\ln x-\mu)^2-2\sigma^2 n\ln x\right)\right]=$$

$$\exp\left[-\frac{1}{2\sigma^2}\left(\ln^2 x-2\mu\ln x+\mu^2-2n\sigma^2\ln x\right)\right]=\exp\left[-\frac{1}{2\sigma^2}\left(\ln^2 x-2\left(\mu+n\sigma^2\right)\ln x+\mu^2\right)\right]=$$

$$\exp\left[-\frac{1}{2\sigma^2}\left(\ln^2 x-2\left(\mu+n\sigma^2\right)\ln x+\left(\mu+n\sigma^2\right)^2-\left(\mu+n\sigma^2\right)^2+\mu^2\right)\right]=$$

$$\exp\left[-\frac{1}{2\sigma^2}\left(\left(\ln x-\mu-n\sigma^2\right)^2-2n\mu\sigma^2+n^2\sigma^4\right)\right]=$$

$$\exp\left[-\frac{1}{2\sigma^2}\left(\left(\ln x-\mu-n\sigma^2\right)^2-2n\sigma^2\left(\mu+0.5n\sigma^2\right)\right)\right]=$$

$$\exp\left[n\left(\mu+0.5n\sigma^2\right)\right]\exp\left[-\frac{1}{2\sigma^2}\left(\left(\ln x-\mu-n\sigma^2\right)^2\right)\right]$$

Using the last equation, formula (C2) can be written as:

$$E\left(x^n \mid a\le x\le b\right)=$$

$$\left[\exp\left(n\mu+0.5n^2\sigma^2\right)\right].\int_{\frac{\ln a-\mu-n\sigma^2}{\sigma}}^{\frac{\ln b-\mu-n\sigma^2}{\sigma}}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}\left(\ln x-\mu-n\sigma^2\right)^2\right]d\left(\frac{\ln x-\mu-n\sigma^2}{\sigma}\right)=$$

$$\left[\exp\left(n\mu+0.5n^2\sigma^2\right)\right]\left[F\left(\frac{\ln b-\mu}{\sigma}-n\sigma\right)-F\left(\frac{\ln a-\mu}{\sigma}-n\sigma\right)\right] \tag{C3}$$

Where $F(.)$ denotes the standard normal distribution function. From equation (C3) it follows that

$$PR\left(a\le x\le b\right)=\left[F\left(\frac{\ln b-\mu}{\sigma}\right)-F\left(\frac{\ln a-\mu}{\sigma}\right)\right]$$

$$E\left(x\mid a\le x\le b\right)=\left[\exp\left(\mu+0.5\sigma^2\right)\right]\left[F\left(\frac{\ln b-\mu}{\sigma}-\sigma\right)-F\left(\frac{\ln a-\mu}{\sigma}-\sigma\right)\right] \tag{C4}$$

$$E\left(x^2\mid a\le x\le b\right)=\left[\exp\left(2\mu+2\sigma^2\right)\right]\left[F\left(\frac{\ln b-\mu}{\sigma}-2\sigma\right)-F\left(\frac{\ln a-\mu}{\sigma}-2\sigma\right)\right]$$

From (C4) we see that the mathematical expectation of x, $E(x)$ equals:

$$E(x) = \int_0^\infty x g(x) dx = \exp\left(\mu + 0.5\sigma^2\right) \tag{C5}$$

Similarly, the mathematical expectation of x^2, $E(x^2)$, obeys:

$$E(x^2) = \int_0^\infty x^2 g(x) dx = \exp\left(2\mu + 2\sigma^2\right) \tag{C6}$$

To compute values of the standard normal distribution function, we apply the numerical approximation given in Abramowitz and Stegun (1972).

In the calibration of the model, we use data on the coefficient of variation of health care spending. To be able to do this, we first use equation (C4) to derive an expression for the variance of health care spending.

$$\begin{gathered}\sigma_x^2 = E(x^2) - \left(E(x)\right)^2 = \int_0^\infty x^2 g(x) dx - \exp\left(2\mu + \sigma^2\right) = \\ \exp\left(2\mu + 2\sigma^2\right) - \exp\left(2\mu + \sigma^2\right) = \exp\left(2\mu + \sigma^2\right)\left(\exp(\sigma^2) - 1\right) = \left(E(x)\right)^2 \left(\exp(\sigma^2) - 1\right)\end{gathered} \tag{C7}$$

It follows that we may write the coefficient of variation of health care spending as:

$$C_v = \frac{\sigma_x}{E(x)} = \sqrt{\exp\left(\sigma^2\right) - 1} \tag{C8}$$

Appendix D: Calibration of model parameters

The calibration upon data for the privately insured implies that 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 the number of hospital admissions to the number of patients: 8%. Data have been obtained from Statistics Netherlands (2006).

We calculate the values of σ_O and σ_A from the corresponding coefficients of variation of the log-normal distribution functions of medical need γ . For the latter, we use the coefficients of variation of total health care spending as calculated in Van Vliet and Van der Burg (1996).

Finally, the value of $\gamma_{\min}$ 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/).

To obtain values for parameters that cannot be directly computed from observed data, we use an adapted model. This model contains additional equilibrium conditions to guarantee that the values of the unknown parameters are set such that the model meets observed data on a number of key variables. These 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 and outpatient 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.

The first five key variables together determine the values of the five unknown parameters β, δ_O, δ_A, μ_A, μ_O and π_0. For example, as can be seen from equation (9), demand for medical services is a function of δ_O, δ_A, $\mu_A\left(=E(\gamma_A)\right)$ and μ_O. We use total co-payments per patient to compute a value of the co-payment maximum *m* that fits the data.

Given the values of the key variables, calibration proceeds in a number of steps:

1. Set initial values for the demand parameters β, δ_E and δ_A and the copayment maximum *m*. The value of β depends on the value of the CRRA parameter and the net average patient income: real gross income minus health care premiums and copayments. In the first step, we compute β using average real gross income per patient.

2. Run the adapted model; this model includes three loops, that determine: (i) the probability of zero medical need π_0 such that the probability of zero costs meets its observed value; (ii) the value of δ_A such that the insurance effect for the A group meets the value of in table 1 and (iii) the expected medical need ($E(\ln\gamma)/\delta$) for each group such that computed and observed demand coincide. From the medical need and the values of δ_E and δ_A the parameters μ_A and μ_E can be computed. The calibration model also yields values of the income and price elasticities per group, the expected values of $E(\ln\gamma)$ for both groups, the average co-payments per group, the exact value of the CRRA and the exogenous probability of zero costs that does not depend on parameters of the copayment regime.

3. Adjust the values of the parameters β, δ_E and *m*. The value of β directly affects the CRRA; a smaller value of δ_E increases the absolute value of the price elasticity, lowers the income elasticity and the exogenous probability of zero costs. Finally, a higher value of the

copayment maximum *m* increases the absolute value of the price elasticity, the exogenous probability of zero costs and copayments.

4. Return to step 2 until the calibrated parameter values generate the correct values of the key variables.

To obtain a value of the copayment maximum *m* we note that the paper assumes that insurance covers all curative health expenditures. In practice however, medical insurance only partly compensates expenditures related to dentists, physiotherapists and medicines. Therefore, we choose the co-payment maximum *m* such that predicted co-payments meet the data. Actual copayments have been obtained from the Statistics Netherlands Health Accounts (2006), but copayments related to dentists, physiotherapists and medicines correspond to all insured patients, both the privately insured and the former publicly insured. Depending on the allocation between the two groups, values varying from 108 euro and 257 euro per privately insured patient are possible. A proportional allocation yields a value of 164 euro per privately insured patient.

Second, we use the price elasticity of the demand for outpatient care as estimated in Van Vliet (2001): -0.079. The corresponding value of -0.007 for hospital services is not directly applicable, however. Given the minimum expenditures of 2000 euro for patients in the A-group, their expenditures always exceed the coinsurance maximum – which is about 300 euro per patient - so our model calculates the price responsiveness of the demand of the A-group to be zero. To circumvent this problem, we translate the estimate of the price elasticity into an estimated insurance effect. As noted above, the latter is defined as the ratio of the demand of a fully insured patient (without any co-payments) and the demand of the same patient without health insurance. We calibrate the model parameters such as to reproduce this insurance effect.

Table D.1 summarizes the data used to calibrate the model. Table D.2 discusses economic features of the calibrated model. Table D.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 D.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 copayments per patient (euro)			164
Coinsurance rate (%)			100

Table D.2 Validation of model parameters: results

	Group O	Group A	Total
Probability of zero need: π_N (%)			1.8
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 of quadratic non-medical consumption in utility function: β			$1.94\ 10^{-5}$
Parameter of quadratic health care consumption 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: $\gamma_{\min}/\delta$	0.0	5.5	
Average copayments per patient			159
Copayment maximum	300	300	
Income elasticity health care consumption	0.16	0.0	0.15

Table D.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