



CPB Discussion Paper | 239

Why it may hurt to be insured: the effects of capping coinsurance payments

Ed Westerhout
Kees Folmer

Why it may hurt to be insured: the effects of capping coinsurance payments

Abstract:

Most health insurance schemes use some sort of cost sharing to curb the moral hazard that is inherent to insurance. It is common to limit this cost sharing, by applying a deductible or a stop loss, for example. This can be motivated from an insurance perspective: without a cap, coinsurance payments might be unacceptably high for people with high medical costs. This paper shows that introducing a cap on coinsurance payments may actually hurt people with high medical costs. This is not due to moral hazard that comes along with the extra insurance. Instead, it is because the introduction of a cap makes health spending below the cap more price elastic, thereby inducing the health insurer to raise the coinsurance rate.

Keywords: Moral Hazard, Deductibles, Co-Payment Schemes in Health Care, Idiosyncratic Health Shocks.

1 Introduction

Health reform is high on the political agenda. The introduction in the US of the Patient Protection and Affordable Care Act (Obamacare) is a primary example. Other countries are debating reforms as well. There, the focus is not so much on accessibility as in the US, but more on containing costs. One reason is the excessive growth in health spending in the last decades (OECD (2011)), driven primarily by demography, income growth (Hall and Jones (2007)) and medical technology (Chandra and Skinner (2012)). Another is the almost universal prospect of population ageing which may trigger a further increase in medical spending.

One way to reduce health care spending is a more intensive use of cost sharing. Many consider cost sharing an effective way to reduce health care spending; on whether the benefits of cost sharing outweigh the costs, there is considerable less agreement. A higher reliance on coinsurance can take several forms. A higher rate of coinsurance or a higher deductible are obvious examples. A third example is the partial or full exclusion of typical services from insurance.

One of the reasons for the disagreement on the attractiveness of cost sharing is that cost sharing not only reduces the implicit subsidy on health care consumption and thereby the associated welfare loss, but also reduces the scope of insurance and the gains associated with it. Hence, an optimal cost sharing scheme strikes a balance between the loss from moral hazard and the gain from risk sharing. This coincides with the result of formal welfare analysis: a marginal change in the coinsurance rate that corresponds to the optimal cost sharing scheme would imply a change in insurance and a change in the insurance premium, of which the utility effects would offset each other exactly.

However, a scheme that puts an upper limit on coinsurance payments must yield even higher expected utility than a linear scheme that subjects all medical spending to the same rate of coinsurance. Indeed, the former scheme features two policy instruments that can be chosen optimally - the coinsurance rate and the coinsurance maximum - whereas the linear scheme features only one policy instrument.

This paper analyses the properties of two schemes: one without a cap on coinsurance payments (a linear insurance scheme) and one with a cap. It compares the two schemes on their economic and welfare effects in a variety of settings. In particular does the paper focus on the implications of imposing a cap on coinsurance payments upon the consumers of inpatient services and the consumers of outpatient services. The paper adopts a numerical simulation model: as the cap on coinsurance makes the budget constraint of consumers nonlinear, the model is too complex to be solved analytically.

The analysis confirms that the scheme that uses a cap features a higher level of expected utility than the linear scheme. Although the implications of the reform for consumer utility may be modest at the aggregate level, they are not at a lower level of aggregation. Indeed, the reform may imply huge transfers between the consumers of inpatient and outpatient services. Moreover, the paper achieves the surprising result that the consumers of inpatient services, who are supposed to obtain more insurance by capping coinsurance payments, may actually lose from the reform. A fraction of the population of inpatient services consumers pays less than the maximum of coinsurance payments and will be subject to higher coinsurance payments if the introduction of a cap is accompanied with an increase of the coinsurance rate. If this element is strong enough, the cap on coinsurance may lower the welfare of inpatient services consumers.

There is some earlier literature on optimal coinsurance schemes. Feldstein (1973), Arrow (1976) and Feldman and Dowd (1991) assumed that the budget constraints of households are linear. Keeler *et al.* (1977), Ellis (1986) and Manning and Marquis (1996) did account for the nonlinearity of the consumer's budget constraint, but did not explore the implications for the optimal cost sharing scheme.

A paper that is close to ours is Kowalski (2011). Like our paper, this paper analyzes the value of moral hazard and that of insurance with one and the same model, thereby avoiding possible inconsistencies between the welfare loss from moral hazard and the welfare gain from risk protection. In this, the two papers differ from Feldstein (1973), Feldman and Dowd (1991), Finkelstein and McKnight (2008), Feldstein and Gruber (1995) and Engelhardt and Gruber

(2010), who assess these two aspects separately. Different from our paper, Kowalski (2011) does not explore the optimal coinsurance scheme, however.

The structure of our paper is as follows. Section 2 sets up a model of health care demand with a kinked budget constraint. Section 3 explains how we use the model to derive optimal insurance policies. Section 4 describes the calibration of our model on data for the Netherlands. Section 5 presents our results. Section 6 contains concluding remarks.

2 The simulation model

2.1 The distribution of medical need

This paper characterizes the optimal coinsurance scheme in the class of schemes that apply a uniform rate of coinsurance up to some maximum. The optimal scheme maximizes expected utility of a representative consumer who is risk-averse, price-responsive and aware of the nonlinearity of his budget constraint. Expected utility integrates the utility levels that correspond to different exogenous shocks to the health status of the consumer. Maximization is done by choosing optimally the coinsurance rate and maximum of coinsurance payments.

Central in our model is individual health status, an exogenous variable that varies between consumers. The distribution function of the health status variable has to meet some stylized facts. First, in any year there are many people who do not consume any health services at all. Second, there is a large group of people who consume health services, but do not use any inpatient care. The size of the remaining group, people who consume both outpatient and inpatient services is much smaller. Finally, health expenditure of the two latter population groups can reasonably well be described using lognormal distribution functions (Feenberg and Skinner (1994), French and Jones (2004)). The parameters of these two distributions are very different: the distribution function for the health expenditure of those who consume inpatient services features a higher minimum amount of expenditure, a higher mean and is more skewed to the right than the distribution function for the health expenditure of those who consume outpatient services only.

Based on these observations, we argue that a mixture of three distributions, namely two lognormal distributions and a mass point at value zero, yields a tractable and sufficiently realistic formulation of the heterogeneity of medical need. The four-part approach that is often used in empirical analysis (Duan *et al.* (1983)) combines these findings. Our approach is based upon this four-part model, but relates it to health status, characterized by a parameter γ , rather than to health care expenditure. In particular, we assume that a population fraction π_0 does not need any health care ($\gamma = 0$). Further, a population fraction $(1 - \pi_0)\pi_o$ consumes outpatient (O) services and a population fraction $(1 - \pi_0)\pi_i = (1 - \pi_0)(1 - \pi_o)$ consumes inpatient (I) services. The medical needs of patients for O services and I services are described by two lognormal distribution functions that differ on two accounts. First, there is a minimum medical need, γ_{\min} , that is zero in the case of outpatient services and positive in the case of inpatient services. Second, the means and variances of the distributions that correspond to the need for O and I services in excess of the corresponding minimal need, $\gamma - \gamma_{\min}$, are different.

2.2 Consumer behaviour

We assume that the consumer has partial information on his health status. In particular, he knows whether he belongs to the group with zero health spending, the O group or the I group. He does not know his health status. The consumer chooses his health care consumption after his health status has been revealed. This and the next section discuss first the optimization problem in general. Thereafter, the differences between the O case and the I case are explained.

We adopt a quadratic utility function, where u denotes the patient's direct utility, z denotes the consumption of health care and c denotes the consumption of other, non-medical services. Parameters and variables differ between the O case and I case, but, for brevity, we refrain from attaching explicit indices.

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

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

Here, y is the average gross income per patient. The constraint on the parameter β ensures that the marginal utility of non-medical consumption is positive everywhere.

The quadratic form of the utility function has several attractive properties. First, it ensures that the demand for medical services is finite even in the case of a zero out-of-pocket price: A utility function with positive marginal utility of medical consumption everywhere does not yield an interior solution in case the out-of-pocket price is zero. Second, the quadratic form implies that the price elasticity of health care demand is decreasing in health status and increasing in the co-payment rate, which is backed by empirical evidence (Wedig (1988), Phelps and Newhouse (1974)). By implication, average spending on hospital services is less price elastic than average spending on outpatient services, as is the case in the data (Newhouse *et al.* (1993), Chandra *et al.* (2012)). Moreover, the combination of a quadratic form and lognormality of the distribution functions is analytically convenient. It produces analytical solutions for the expected values of functions of variables. Hence, we do not have to rely on stochastic simulation to find the optimal coinsurance scheme, which saves time and avoids simulation errors due to insufficient sample size.

The quadratic form is generally regarded as unappealing because it has the unrealistic implication that risk aversion is increasing in income. This argument has little weight here since our analysis does not account for income heterogeneity. We admit that different coinsurance schemes differ in terms of premiums and coinsurance payments and thus in terms of disposable income. However, the implied differences in disposable income are too small to exert important effects upon risk aversion.

As discussed above, the parameter γ is different for different states of health. This reflects patient heterogeneity in terms of the need of health care (the marginal utility of health care consumption equals $\gamma - \delta z$). This may be interpreted as reflecting a state-dependent health production function in which 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.

The rate of coinsurance is denoted as b . The 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 (different for the O and I group) so that bt measures the out-of-pocket price of medical services. On account of the maximum to coinsurance payments, the budget constraint of the consumer is nonlinear:

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

where y_p is defined as $y - p$, y denotes gross income and p denotes health insurance premiums.

The consumer acts rationally. Hence, he maximizes (1), subject to (2), given that he knows the value of the parameter γ that reflects his health status. This problem is non-standard due to the endogeneity of the kink in the budget constraint, *i.e.* the consumer's choice for z defines the out-of-pocket price of his marginal unit of medical consumption. Indeed, the optimization problem has to be solved in another way.

2.3 The demand for health care

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 step 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:

$$\begin{aligned}
z &= 0 & 0 \leq \gamma \leq \gamma_0 \\
z &= \frac{\gamma - bt(1 - \beta y_p)}{\delta + \beta(bt)^2} & \gamma_0 \leq \gamma \leq \gamma_1 \\
z &= \frac{\gamma}{\delta} & \gamma_1 \leq \gamma
\end{aligned} \tag{3}$$

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

$$\begin{aligned}
\gamma_0 &= bt(1 - \beta y_p) \\
\gamma_1 &= \frac{-\delta(1 - \beta y_p)}{\beta(bt)} + \frac{\delta}{\beta(bt)^2} \sqrt{\Omega} \\
\Omega &= (bt)^2 (1 - \beta y_p)^2 + 2(bt)^2 \beta \left[\frac{1}{2} \beta m^2 + m(1 - \beta y_p) + \frac{(bt)^2}{\delta} (1 - \beta(y_p - m))^2 \right]
\end{aligned} \tag{4}$$

Equation (3) implicitly assumes that $\gamma_1 > \gamma_0$. It can be derived that this requires the coinsurance maximum m to exceed some minimum value. In all the simulations reported in this paper, the coinsurance maximum exceeds this minimum value.

Equation (3) demonstrates that health care demand is a piecewise linear function of γ . The economics behind this equation can be illustrated by following what happens when γ increases from zero to infinity. A state of perfect health ($\gamma = 0$) obviously implies zero demand. Minor illnesses, defined by $0 < \gamma \leq \gamma_0$, produce zero demand too. For this type of 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 $\gamma \geq \gamma_0$, health care demand is positive and increasing with the severity of the 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 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 increasing in 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 as the three equations for z :

$$\begin{aligned}
 v &= y_p - 0.5\beta y_p^2 & 0 \leq \gamma \leq \gamma_0 \\
 v &= \frac{\delta y_p + 0.5(bt)^2 - 0.5\beta\delta y_p^2 + 0.5\gamma^2 - \gamma(bt)(1 - \beta y_p)}{\delta + \beta(bt)^2} & \gamma_0 \leq \gamma \leq \gamma_1 \\
 v &= \frac{\delta(y_p - m) - 0.5\beta\delta(y_p - m)^2 + 0.5\gamma^2}{\delta} & \gamma_1 \leq \gamma
 \end{aligned} \tag{5}$$

As discussed above, we distinguish between O care and I care. In case of O care, the minimum value of γ , γ_{\min} , equals zero. Equations (3) to (5) hold unambiguously. In case of I care, $\gamma_{\min} > 0$. Now, two cases are possible: (i) $\gamma_{\min} < \gamma_0$. Equations (3) to (5) apply, except that the first segment is $\gamma_{\min} \leq \gamma \leq \gamma_0$, rather than $0 \leq \gamma \leq \gamma_0$. (ii) $\gamma_{\min} > \gamma_0$. The first segment of equations (3) to (5) does not apply and the second segment is $\gamma_{\min} \leq \gamma \leq \gamma_1$, rather than $\gamma_0 \leq \gamma \leq \gamma_1$.

In addition, we distinguish between bounded and unbounded coinsurance schemes. Bounded schemes feature a finite coinsurance maximum. In this case, equations (3) to (5) 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 (4) that, in this case, γ_1 is infinite and, thus, the third segments of equations (3) and (5) do not apply.

3 The optimal coinsurance scheme

For both inpatient services and outpatient services, we integrate the levels of consumer utility for all values of γ to arrive at an expression for expected utility. Let $G(\cdot)$ denote the distribution function of γ_i , ($i = O, I$). Then we obtain:

$$\begin{aligned}
 V_i = & G(\gamma_{i,0}) \left(y_p - \frac{1}{2} \beta y_p^2 \right) + \left(G(\gamma_{i,1}) - G(\gamma_{i,0}) \right) E(v_i | \gamma_{i,0} \leq \gamma_i \leq \gamma_{i,1}) \\
 & + \left(1 - G(\gamma_{i,1}) \right) E(v_i | \gamma_i \geq \gamma_{i,1}) \quad i = O, I
 \end{aligned} \tag{6}$$

The conditional expectation variables $E(v | \cdot)$ in equation (6) can be derived from the expressions in equation (5).

Expected utility can now be written as follows:

$$\begin{aligned}
 V = & \pi_0 V_0 + (1 - \pi_0) \pi_O V_O + (1 - \pi_0) \pi_I V_I \\
 = & \pi_0 \left(y_p - \frac{1}{2} \beta y_p^2 \right) + (1 - \pi_0) \pi_O V_O + (1 - \pi_0) \pi_I V_I
 \end{aligned} \tag{7}$$

We define social welfare as aggregate expected utility. This weighs the expected utility levels in case of zero health care, outpatient care and inpatient care. The expected utility levels of the latter two cases are defined in equation (6).

The social welfare function reflects our assumptions on the information set of consumers. Consumers know whether they belong to the group that does not consume health care, to the O group that consumes outpatient services only or to the I group that consumes also inpatient services. They do not know ex ante how much health care they will be consuming as they do not know the size of their health status γ . Insurers, on the other hand, do not observe anything about the health status of the insured. Hence, they cannot charge different premiums to different types of consumers. Obviously, these assumptions are stylized and, taken literally, unrealistic.

However, they are closer to reality than the alternative assumptions of full information or complete uncertainty.

It is useful to point at three other assumptions as well. The first is that we limit the benefits from health insurance to the gain from the avoidance of financial risk. As argued by Nyman (1999), a second benefit from health insurance is that insurance can make types of health care services available that otherwise would not be affordable. Hence, the benefits of health insurance may be underestimated in our model, in particular for persons in the I group. Whether this is very relevant for our analysis of reforms that do not change the degree of insurance, but rather shift insurance between different health shocks, is unclear. An indication of the relevance of this issue may be found in exploring the effects of a higher degree of risk aversion, which we will do in the sensitivity analysis.

The second assumption that deserves discussion is that of the rationality of the health consumer. We stick to the assumption that the household is a rational decision maker who uses all the information that is available. There is ample evidence that this standard assumption of economics does not hold in reality, however. Irrational behaviour may be particularly relevant in the health sector where information imperfections are numerous. There is little agreement on what model of the decision-making process would be a better alternative to the standard rational decision maker model, however.

The third point concerns the role of suppliers of medical care. The concept of supplier-induced demand suggests that supply factors also play a role in the decision process of the health consumer. Including these factors into our model could qualify our results. Modelling the interaction between patients and suppliers of medical care services seems so complex however that it would take a separate paper.

Equation (7) expresses expected utility in three ways to the policy variables b and m . First, the conditional expectation variables $E(v | \cdot)$ relate to the policy variables (equation (5)). Second, the threshold values γ_0 and γ_1 relate to the policy variables (equation (4)). Thirdly, after-premium income, y_p , relates to the policy variables through health insurance premiums (recall $y_p = y - p$).

4 Calibration of the model

Our simulation procedure is to calibrate the model upon data for that part of the population in the Netherlands that was privately insured before the 2006 health insurance reform. This determines the parameter configuration that we use for our benchmark simulation. Other simulations assume (much) lower and higher values for the parameters of our model. This allows us to explore the robustness of our results and underlines that our results are not specific to the parameter configuration that we employed for the benchmark simulation.

The nine parameters that we have to quantify are the probability of zero need for medical services, π_0 , the probability of need for inpatient services, π_I , the parameters that describe the moments of the two lognormal distribution functions for the γ parameter, *i.e.* μ_0 , μ_I , σ_0 and σ_I , and the remaining parameters of the household utility function, *i.e.* β , δ_0 and δ_I . The nine variables on which we calibrate our model are the frequency of zero medical spending, the frequency of inpatient spending (defined as including at least one hospital admission), the coefficients of variation of the distributions of health care expenditure of patients in the O and I group, average health care demand on inpatient and outpatient spending, the coefficient of relative risk aversion for non-medical products, the price elasticity of health care demand for O services and the insurance effect for I-services.

The services covered are pharmaceuticals and services delivered by general practitioners, dentists, physiotherapists and hospitals (inpatient and outpatient care). We use the estimates in Van Vliet and Van der Burg (1996) of the coefficients of variation of the lognormal distribution functions for health care expenditure of people with and without consumption of inpatient services in order to pin down the standard deviations of the distributions of γ_E and $\gamma_I - \gamma_{\min}$.

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. We take the expenditures corresponding to γ_{\min} equal to 2000 euro (in prices of 2010).

We use the price elasticity of the demand for outpatient care as estimated in Van Vliet (2001): -0.079. We cannot use Van Vliet's (2001) price elasticity of -0.007 for hospital services for calibration, however. Given the minimum expenditures of 2000 euro for patients in the I-group, their expenditures always exceed the coinsurance maximum, so our model calculates the price responsiveness of the I-group to be zero. To circumvent this problem, we translate the price elasticity estimate into an insurance effect, 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 and calibrate the model such as to reproduce this insurance effect.

Newhouse *et al.* (1993) reports price elasticities as obtained in the RAND Health Insurance Experiment (HIE). These are in the range of (-0.1, -0.2) and are higher in absolute value than the results of Van Vliet (2001). Section 5 will explore how sensitive are our results with respect to the assumption on the price elasticity of health care demand.

Finally, for the coefficient of relative risk aversion (CRRA) for non-medical products, we adopt a value of 2.

Table 1 Validation of model parameters: data

	Group O	Group I	Total
Probability of zero expenditure (%)			22.1
Probability of positive expenditure on outpatient and inpatient services (%)			8.0
Insurance effect I-group		1.03	
Price elasticity of health care demand	- 0.079		
CRRA, non-health products			2.0
Coefficient of variation health care costs	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
Co-payment rate (%)			100

Table 2 Validation of model parameters: results

	Group O	Group I	Total
Probability of zero need: π_0 (%)			1.8
Relative size of the E-group: π_E (%)			90.2
relative size of the I group: π_I (%)			8.0
Parameter of quadratic non-medical consumption in utility function: β			$1.94 \cdot 10^{-5}$
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: γ_{\min} / δ	0.0	5.5	

Tables 1 to 3 summarize the calibration of our model. What stands out is the skewed distribution of need and demand. If a person spends on health care, there is a chance of only about 1 to 10 that this spending includes inpatient hospital services. However, the average volume of health care consumption of I services is almost double the size of O services and the price of the former is about ten times as large as the price of outpatient services.

About a quarter of the population does not consume any health care. Of this, only 2.1% has zero medical need. For the largest part, people that choose to have zero spending are people with positive need for outpatient services that is so small that the benefits from medical intervention are less than the costs involved. Table 3 summarizes the resulting population fractions in the O and I group. Interesting is what these figures imply for the incidence of co-payments. About 20% of the population has zero co-payments and about 40% pays the maximum (the deductible); hence about 40% of the population in the dataset that we use for calibration co-pay in between zero and the maximum.

With respect to outpatient care, the volume of need is estimated to be 15.9 units per person, some 7% larger than demand (14.5). With respect to inpatient care, need and demand coincide at the level of 24.9, in line with our result that in the calibrated economy no consumers of I services co-pay in between zero and the deductible.

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

The calibration of the model implies an average income elasticity of health care demand of 0.50. This combines an income elasticity of 0.57 for outpatient services and of 0.0 for inpatient services. The reason for the zero income elasticity of the I-group is that there are no patients in

this group who have health care expenditure below the maximum of co-payments. Although our estimate of the income elasticity is higher than estimates based on micro data and lower than estimates based on macro data, it corresponds to Getzen (2000) who reports values ranging from 0.5 to 0.9. Our sensitivity analysis includes simulations with much smaller and larger values for the income elasticity of health care demand.

5 Numerical simulations

5.1 The optimal unbounded coinsurance scheme

If there is no bound to coinsurance payments, what is then the optimal rate of coinsurance, *i.e.* the rate of coinsurance that maximizes expected household utility? The model is too complex to explore this question analytically. We therefore adopt a numerical approach. In particular, we apply a grid search procedure in which we let the coinsurance rate run from 0% to 100% in steps of 5 percentage points. Our calculations do put a cap on coinsurance payments, namely at the 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 4 displays our results. Its first column shows that the optimal coinsurance rate is 30%. This strikes a balance between the benefits from insurance, *i.e.* smoothing of non-medical and medical consumption across states, and the cost of insurance, *i.e.* a high insurance premium due to the moral hazard in medical consumption. Indeed, based on the smoothing argument, the coinsurance rate should be as low as possible, *i.e.* zero; based on the moral hazard argument, it should be as high as possible, *i.e.* one.

Table 4 Sensitivity analysis optimal unbounded coinsurance scheme

	$\hat{b}(\%)$	\tilde{y}^{ZI}	\tilde{y}^{FI}
Benchmark (BM)	30	99	21
$\beta_i = 0.50\beta_{i,BM}$	55	30	58
$\beta_i = 1.50\beta_{i,BM}$	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
$y = 0.50y_{BM}$	20	211	14
$y = 1.50y_{BM}$	40	59	26
$\pi_E = 0.95\pi_{E,BM}$	30	148	24
$\pi_E = 1.05\pi_{E,BM}$	10	46	19

$i = O, I$

This result depends heavily upon the assumed parameter values. Table 4 also shows the results for cases that feature much smaller or larger values for the degree of risk aversion, the price elasticity of the demand for health care, the standard deviation of the health status, income and the weights of inpatient and outpatient care in total health care. For these cases, the optimal coinsurance rate lies between 10% and 55%. Our results are in line with those by Manning and Marquis (1996), who calculate the optimal coinsurance rate to lie in the range from 40 to 50%. However, our results seem to differ from the result by Feldstein (1973) that increasing the coinsurance rate from 33% to 67% would be welfare-increasing.

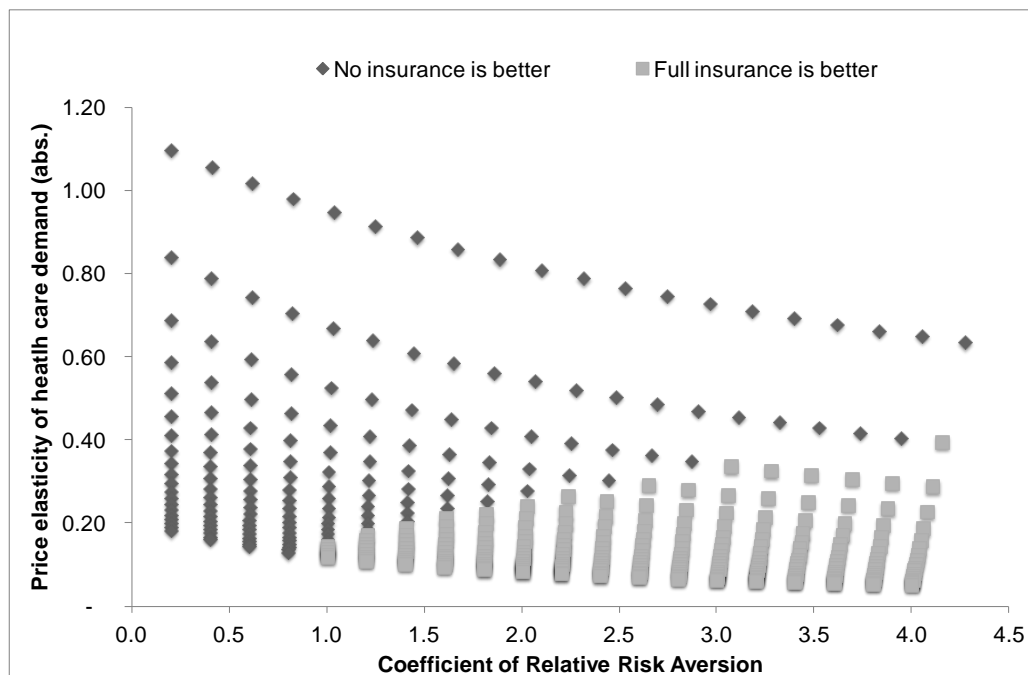
A further interesting question is how much better the optimal scheme is than the polar schemes of zero insurance and full insurance. The second and third columns of Table 4 display the compensating variations of zero insurance and full insurance. We define the compensating variation of an alternative 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. It is a hypothetical

concept; actual compensation does not occur. Formally, the compensating variation of alternative scheme A, \tilde{y}^A , is defined by the equality $V(\hat{b}, \hat{m}, y - \tilde{y}^A) = V(b^A, m^A, y)$, where \hat{b} and \hat{m} denote the coinsurance rate and maximum of the optimal scheme and b^A and m^A denote the counterparts of alternative scheme A. 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. According to this calculation, full insurance is thus superior to no insurance. This result is opposite to the results of Feldstein (1973), Feldman and Dowd (1991) and Manning and Marquis (1996). This result that we achieve for the benchmark case is far from robust, though. The simulations in Table 4 show that full insurance can dominate zero insurance, but that the opposite may be true as well.

A different way to demonstrate that both full insurance and zero insurance can come out better is to choose different combinations of the coefficient of relative risk aversion (CRRA) and the (absolute value of) price elasticity of health care demand. Figure 1 does so (with both the CRRA and the price elasticity measured at the mean of health care demand). For high values for the price elasticity of demand, no insurance dominates zero insurance. For low values, the opposite holds true. Furthermore, for intermediate values for the price elasticity of demand, no insurance is better than full insurance for low values for the CRRA, whereas the opposite holds true for high values for the CRRA.

Figure 1 No insurance versus full insurance



How does optimal insurance compare with zero insurance and full insurance at the level of the O group and I group? The compensating variations corresponding to zero insurance are -342 euro for the O group and 4,328 euro for the I group. This result is peculiar for two reasons. First, the signs of the welfare effects are different. That is, the consumers of inpatient services have a strong preference for optimal insurance rather than zero insurance, but the consumers of outpatient services prefer not to be insured. Second, relative to that for the whole population, the welfare effects are sizeable.

Having seen the results for the zero insurance scheme, the results for the scheme of full insurance are not surprising. The consumers of inpatient services are 1,815 euro better off in the scheme of full insurance; the consumers of outpatient services are 186 euro worse off. Again, the aggregate welfare loss from a suboptimal scheme, in this case full insurance, hides sizeable differences between I care consumers and O care consumers.

5.2 The optimal bounded coinsurance scheme

As said, the optimal unbounded scheme cannot be optimal as it does not exploit the differences across states of nature in the price elasticity of health care demand. Indeed, the price elasticity of health care demand differs for different health status levels: the better is health status (the lower is γ), the more price elastic is demand. Ideally, if the insurer could apply a continuum of coinsurance rates, he would choose for a scheme in which the coinsurance rate is an increasing function of the price elasticity of demand (Blomqvist (1997)). An insurance scheme that features only two coinsurance rates, one positive and the other zero, will be worse than this ideal scheme, but will improve upon the linear scheme as it can mimic more closely the ideal scheme.

What are the implications of choosing optimally the maximum of coinsurance payments? This reduces coinsurance payments for the states of very bad health and allows the insurer to reduce the health insurance premium. If the insurer, as before, also chooses the optimal rate of coinsurance, the introduction of a maximum 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 optimize with respect to coinsurance rate and coinsurance maximum, *i.e.* we explore which combination of coinsurance rate and maximum achieves the highest level of expected utility. Our grid search procedure lets the coinsurance rate run from 0% to 100% in steps of 5 percentage points and the maximum of coinsurance payments from a minimum value of 100 euro to 50,000 euro in steps of 50 euro. Actually, the minimum and maximum values that are used are irrelevant for our results; the coinsurance maxima in our simulations deviate strongly from both this minimum and maximum. Table 5 summarizes our results.

The coinsurance maximum drops to a level of 3,250 euro and the optimal coinsurance rate increases to 55%. On balance, coinsurance payments increase, by about 40%. Aggregate health expenditure and health insurance premiums fall. The increase in coinsurance payments differs between the groups of outpatient and inpatient care consumers. Coinsurance payments in the

states of outpatient care increase on account of the higher coinsurance rate. Surprisingly, at least at first sight, coinsurance payments by inpatient care consumers increase as well! Indeed, a sizeable fraction of the I group has below maximum coinsurance payments and for these people, coinsurance payments increase. This effect of the increase in the coinsurance rate is so large, that it dominates that of capping coinsurance.

Table 5 Effects of reforming the coinsurance scheme			
	<i>A</i>	<i>B</i>	$(B - A) / A (\%)$
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: I 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, I group	2,009.03	2,362.99	17.62
Insurance premium	659.60	529.57	-19.71

The compensating variation of the welfare gain that corresponds to capping coinsurance is very modest: 9 euro only. This is as expected. The corresponding measures of the welfare changes corresponding to schemes of zero insurance and full insurance equal 99 euro and 21 euro respectively, and these schemes deviate much more from the optimal linear scheme. Moreover, the aggregate welfare gain hides substantial transfers between different groups of consumers. Those who consume inpatient services *lose* 201 euro, whereas the consumers of outpatient services gain 29 euro. That the inpatient services consumers lose is counterintuitive. We will return to this issue in the next section.

Table 6 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. This confirms the result from the benchmark case that the introduction of a maximum to coinsurance payments entails an increase of the coinsurance rate and a welfare gain that is modest at the aggregate level. Our result that the consumers of inpatient services lose from maximizing coinsurance payments is not confirmed however. In 5 out of 10 cases (see Table 5), the consumers of inpatient services gain.

Table 6 Sensitivity analysis on optimal bounded coinsurance scheme

	$\hat{b}(\%)$		\hat{m}	\tilde{y}^B	\tilde{y}_O^B	\tilde{y}_I^B
	A	B				
Benchmark (BM)	30	55	3,250	9	29	-201
$\beta_i = 0.50\beta_{i,BM}$	55	70	7,100	8	8	0
$\beta_i = 1.50\beta_{i,BM}$	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
$y = 0.50y_{BM}$	20	90	1,050	13	-53	706
$y = 1.50y_{BM}$	40	60	4,300	8	13	-44
$\pi_E = 0.95\pi_{E,BM}$	30	45	3,500	10	27	-98
$\pi_E = 1.05\pi_{E,BM}$	40	90	1,100	10	-65	1961

$i = O, I$

This can be explained from the maximum 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 assumed in the benchmark case. This corresponds to a relatively high degree of risk aversion. Quantitatively, the results differ strongly from those in the benchmark case. On the one hand, the maximum to coinsurance payments is now 1,100 euro rather than 3,250 euro. In particular, the consumers of outpatient care benefit from this lower maximum. They now benefit 641 euro in compensating variation terms. On the other hand, the increase in the coinsurance rate is 65

percentage points rather than 25 percentage points as in the benchmark case. Particularly, the consumers of outpatient services are hurt by the implied increase in coinsurance payments. They now lose 48 euro in compensating variation terms due to the reform of the insurance scheme.

The counterintuitive result hinges upon the relationship between the price elasticity of health care demand, in absolute terms, and health status. If a deterioration of health status would go along with an increase in the price elasticity of demand, the reform would induce the insurer to reduce the coinsurance rate. The population that consumes inpatient services would definitely gain from the reform. Such a case cannot be mimicked by our model, however. In addition, the assumed behavior of the price elasticity of demand would conflict with available empirical evidence.

5.3 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 6, the welfare loss declines from 9 to 4 euro). This is about half the welfare gain of the reform. The consumers of inpatient services gain 348 euro, the consumers of outpatient services lose 28 euro. Capping coinsurance payments without raising the coinsurance rate benefits consumers with high levels of spending. These consumers are concentrated in the I population. 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, the average outpatient services consumer 57 euro and the

average inpatient services consumer -549 euro. The consumers of inpatient services lose as a large part of this population has spending that is higher than the average spending of consumers of outpatient services but lower than the 3,250 euro maximum of coinsurance payments.

Table 7 Decomposing the reform of the insurance scheme

	$\hat{b}(\%)$	\hat{m}	\tilde{y}^B	\tilde{y}_o^B	\tilde{y}_r^B
Optimal unbounded scheme	30	50,000	9	29	-201
Hybrid scheme	30	3,250	4	57	-549
Optimal bounded scheme	55	3,250			

6 Concluding remarks

The results of our paper indicate that reforms of the co-insurance scheme may have small efficiency effects at the aggregate level, but large effects at a lower aggregation level. Political-economy aspects may therefore make it difficult to get reforms approved. We must admit that many of the numerical results are far from robust. Whether full insurance is superior or inferior to zero insurance, for example, depends heavily upon the assumed values of risk aversion and the price elasticity of health care demand. Similarly, the size of the welfare gain from the health insurance reform and the distribution of it over consumers of inpatient and outpatient services differs strongly between different simulations. Two more aspects are worth noting.

First, in practice, coinsurance often takes the form of a deductible. This paper suggests that a deductible cannot be part of an optimal scheme, whether it caps coinsurance payments or not. It cannot be more than suggestive as our analysis does not account for administration costs and, seemingly, deductibles outperform other coinsurance schemes on administration costs. Administration costs may also be the reason that deductibles are common in reality, although we think that other factors play a role as well here. In particular, we think that it may be difficult to explain to the insured that more complex co-insurance schemes are more preferable than deductible schemes and that this may withhold insurers to introduce them. The gap between the optimal coinsurance rate in our calculations and the 100% rate of a deductible is

quite large in many cases, however. We conclude from this that it would be useful to reconsider the use of deductibles in health insurance.

Second, the health insurance schemes that we have analysed are uniform, *i.e.* apply to all insured. A scheme that would differentiate the parameters of the coinsurance scheme between people that consume different types of services may yield higher utility. Necessary condition is that the insurer knows to which group each consumers belongs, such as to avoid switching by consumers to the most attractive insurance scheme. Analysis of this question is beyond the scope of this paper, but certainly worth considering.

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 Erik Schut for valuable comments on an earlier version of this paper. Any remaining errors are the responsibility of the authors.

References

ARROW, K.J., 1976. The Role of Health Insurance in the Health Services Sector, in: Richard N. Rosett (ed.), Universities-National Bureau Conference Series no. 27, NBER, New York.

BLOMQVIST, A., 1997. Optimal Non-Linear Health Insurance, *Journal of Health Economics* **16**, 303-21.

CHANDRA, A., J. GRUBER and R. McKNIGHT, 2012. The Impact of Patient Cost-Sharing on the Poor: Evidence from Massachusetts, NBER Working Paper 18023.

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

CUTLER, D.M. and M. McCLELLAN, 2001. Is Technological Change in Medicine Worth it?, *Health Affairs* **20**, 11-29.

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.

DUAN, N., W.G. MANNING, C.N. MORRIS and J.P. NEWHOUSE, 1983. A Comparison of Alternative Models for the Demand for Medical Care, *Journal of Business and Economic Statistics* **1**, 115-126.

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

ENGELHARDT, G.V. and J. GRUBER, 2010. Medicare Part D and the Financial Protection of the Elderly, NBER Working paper 16155.

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. 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.

FRENCH, E. and J.B. JONES, 2004. On the Distribution and Dynamics of Health Costs, *Journal of Applied Econometrics* **19**, 705-721.

GETZEN, T.C, 2000. Health care is an individual necessity and a national luxury: applying multilevel decision models to the analysis of health care expenditures, *Journal of Health Economics* **19**, 259-270.

- GOLDMAN, D. and T.J. PHILIPSON, 2007. Integrated Insurance Design in the Presence of Multiple Medical Technologies, *American Economic Review* **97**, 427 - 32.
- 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., 2011. Estimating the Tradeoff Between Risk Protection and Moral Hazard with a Nonlinear Budget Set Model of Health Insurance, mimeo.
- 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.
- NEWHOUSE, J.P. and the Insurance Experiment Group, 1993. *Free for All? Lessons from the Health Insurance Experiment*, Harvard University Press, Cambridge.
- NYMAN, J.A., 1999. The Value of Health Insurance: The Access Motive, *Journal of Health Economics* **18**, 141-152.
- OECD, 2011. Health at a Glance 2011, OECD indicators, doi: 10.1787/19991312.

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.

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.



Publisher:

CPB Netherlands Bureau for Economic Policy Analysis
P.O. Box 80510 | 2508 GM The Hague
T (070) 3383 380

March 2013 | ISBN 978-90-5833-593-7