ARE DIFFERENTIAL CO-PAYMENT RATES APPROPRIATE IN THE HEALTH SECTOR?

by

Louis EECCHHOUDT
Catholic Faculties of Mons and Lille
Associate Member of CORE

Maurice MARCHAND
U.C.L. and CORE

Pierre PESTIEAU
University of Liège and CORE

Gwenaël PIASER
CORE

This is a very preliminary version. It should not be quoted without the authors' permission.
In countries which rely mainly upon public funds to finance health care, costs are often contained by identical increases in the co-payment rates applied to a variety of medical activities. For instance, the co-payment rate for visits to specialists (in ophthalmology, cardiology, internal medicine and so on) may be uniformly raised from e.g. 20 to 30 per-cent in the hope of containing the increase in health expenditures for the social security system.

While obvious administrative reasons may justify uniform co-payment rates (or uniform increases in them), the purpose of this paper is to show that there are also good economic reasons to differentiate co-payment rates according to the nature of the medical activity.

To illustrate our point of view we consider a group of patients who face the risk of suffering from a single disease during one period. To fight the disease two mutually exclusive medical strategies are available:

- either patients are regularly treated from the very beginning of the period when one knows only the probability they will get the disease. We refer to this strategy as “long term treatment” (L.T.T.) ;
- or nothing is done pro-actively and patients are treated “in emergency” when a medical accident reveals during the period they have the disease. This is the “emergency strategy” (E.S.).

Our objective in this paper is to show that from an economic point of view there are good reasons to differentiate the public subsidies granted to each medical strategy.

To reach this objective, the paper is organized as follows. In section 1 we present the model and its notation. Section 2 is devoted to individual choices that are influenced both by the medical conditions and the co-payment rates applied to each medical strategy as well as the level of taxes.

In section 3 we compare the first best solution of the problem to a “laissez-faire” one. Section 4 is devoted to a “second-best” approach. It is in this section that we prove our main result about the optimal co-payment rates. We end up with a short conclusion and perspectives for future research.

---

1 They are mutually exclusive for each patient. However globally one strategy may be applied to a group of patients and the other strategy to another group.
SECTION 1: THE MODEL

We consider a population of patients with identical increasing, concave and additive utility functions for wealth \( U(W) \) and for health \( V(H) \). These patients are also endowed with the same initial wealth. The only difference between them is the probability \( \sigma \) of developing a given disease during the period.

At the beginning of the period, nobody knows who will become sick; only the probability that each patient becomes sick during the period is public knowledge. The distribution of the probabilities of disease is represented by a density function \( f(\sigma) \) while the cumulative distribution is denoted \( F(\sigma) \).

If someone becomes sick without receiving treatment his health state is denoted \( H_0 \) while full health (absence of disease and of treatment) implies of course a higher level of \( H_1 \), denoted \( H_2 \) with \( H_2 > H_0 \).

In summary the lottery faced by a patient who does not get any treatment is represented by:

\[
\begin{array}{c}
\sigma \\
\text{U}(W-T)+V(H_0) \\
1-\sigma \\
\text{U}(W-T)+V(H_2)
\end{array}
\]

where \( W-T \) is wealth (W) net of taxes (T).

If at the beginning of the period the patient chooses the long term treatment (L.T.T.) he modifies the lottery he faces. While \( \sigma \) is unchanged by this strategy, the net financial and medical outcomes are modified. The net cost of the strategy for the patient is \( (1-s)C \) where \( C \) is the cost of L.T.T. and where \( (1-s) \) is the co-payment rate (so that \( s \) is the subsidy rate determined by the regulator). As far as its medical efficiency is concerned the long term treatment improves the health state of sick people by \( b \) (with \( H_0 + b < H_2 \)) and it does not deteriorate health for patients who do not become sick.

The other available medical strategy corresponds to emergency medicine. When it is adopted no treatment is initiated at the beginning of the period but if the disease occurs during the period, an emergency treatment is used. Its net cost to the patient is \( (1-s)\tilde{C} \) where \( (1-s) \) is the co-payment rate applied to this treatment. As is natural we assume that \( \tilde{C} \geq C \) and that the emergency treatment is medically less efficient than the long term one. Indeed a sick patient
who receives the emergency treatment (E.T.) benefits from a health state \( H_0 + \tilde{b} \leq H_0 + b \)
where \( \tilde{b} \) is the medical benefit of E.T. for a sick patient.

Of course if we want that sick patients (who are not under L.T.T.) adopt the emergency
treatment, we need:

\[
U(W - T - (1 - \tilde{s})\tilde{C}) + V(H_0 + \tilde{b}) > U(W - T) + V(H_0)
\]

(1.1)

We assume that this condition is always satisfied even when \( \tilde{s} = 0 \).

The choice between the two treatments may be represented by a decision-tree:

![Decision Tree](image)

Before looking at individual choices it is worth pointing out that the E.T. strategy is not
dominated by the L.T.T. one despite the fact that it is medically less efficient and more costly.
Indeed the emergency treatment has an “informational” advantage: since it is applied ex-post
when the true state of the world is revealed it is given only to sick people. On the contrary,
the long-term treatment is used as well for sick people as for healthy ones. As a result its
lower cost per patient \( (C \leq \tilde{C}) \) is compensated for by the fact that is applied to a larger
population.
SECTION 2: INDIVIDUAL CHOICES

At the beginning of the period a patient who is informed only about the probability $\sigma$ that he will develop the disease must decide to enter or not to enter the long-term treatment. Of course the higher $\sigma$ the more attractive to him the long-term treatment. In fact there exists a probability threshold $\hat{\sigma}$ such that the patient is indifferent between the two choices described in the decision tree. Formally $\hat{\sigma}$ is solution of:

$$U(W - T - (1 - s)C) + \hat{\sigma} V(H_0 + b) = \hat{\sigma} U(W - T - (1 - \hat{\sigma})C) + (1 - \hat{\sigma}) U(W - T) + \hat{\sigma} V(H_0 + \tilde{b})$$

(2.1)

Patients for whom $\sigma$ exceed $\hat{\sigma}$ turn to the long term treatment while others “wait” and use the emergency treatment later on in case of need.

It is easy to determine how $\hat{\sigma}$ is influenced by the decisions of the regulator about the variables he controls (s, $\hat{s}$ and T). Implicit differentiation of (2.1) yields:

$$\frac{d\hat{\sigma}}{ds} = \frac{-CU'(B)}{-D} < 0$$

(2.2)

$$\frac{d\hat{\sigma}}{d\hat{s}} = \frac{\hat{\sigma}CU'(\tilde{B})}{-D} > 0$$

(2.3)

$$\frac{d\hat{\sigma}}{dT} = \frac{U'(B) - \hat{\sigma}U'(\tilde{B}) - (1 - \hat{\sigma})U'(A)}{-D} = 0$$

(2.4)

where

$$A = W - T$$
$$B = W - T - (1 - s)C$$
$$\tilde{B} = W - T - (1 - \hat{s})C$$
$$D = U(B) - U(A) - V(H_0 + b) + V(H_0 + \tilde{b}) < 0$$

In accordance with intuition (2.2) and (2.3) indicate that if a treatment is more generously subsidized it will be more often adopted by patients. For instance if emergency medicine is better subsidized, $\hat{\sigma}$ increases so that some patients decide not to use the long-term treatment. Among these patients some of them will become sick and will need the emergency treatment.

The impact of T on $\hat{\sigma}$ is sign ambiguous essentially because the wealth effect is itself sign-ambiguous.
SECTION 3 : THE “FIRST-BEST” AND “LAISSEZ-FAIRE” SOLUTIONS

Before turning in the next section to the optimal selection of $s$, $\bar{s}$ and $T$ by the regulator, we first compare a “first-best” solution to a “laissez-faire” one.

If the regulator adopts a “laissez-faire” (L.F.) approach, he implicitly chooses $s = \bar{s} = T = 0$ and simply observes individual choices. In contrast, in the first-best solution (F.B.) the regulator determines a threshold of $\sigma$ above which patients are forced to take – free of charge – the long-term treatment. Below the threshold patients are denied L.T.T. and are referred to the emergency treatment – again free of charge – in case of need. The program is financed by a uniform tax levied on all patients at the beginning of the period.

These two programs are analyzed in turn and are then compared for a specific case.

3.a. The “laissez-faire” solution (L.F.)

In this case, the choice of individual patients is determined as in eq.(2.1) with $T = s = \bar{s} = 0$. Hence the threshold probability at which patients are indifferent between each possible decision ($\hat{\sigma}_{LF}$) is solution of:

$$\hat{\sigma}_{LF}U(W - \bar{C}) + (1 - \hat{\sigma}_{LF})U(W) - U(W - C) = \hat{\sigma}_{LF}\left[V(H_0 + b) - V(H_0 + \bar{b})\right]$$  \hspace{1cm} (3.1)

Once $\hat{\sigma}_{LF}$ is obtained the proportion of patients who use L.T.T. is simply equal to $1 - F(\hat{\sigma}_{LF})$ while $\int_{0}^{\hat{\sigma}_{LF}} \sigma dF(\sigma)$ patients will need the emergency treatment later on.

Using the concept of the risk premium, (3.1) can also be written as:

$$U(W - \hat{\sigma}_{LF}\bar{C} - \pi) - U(W - C) = \hat{\sigma}_{LF}\left[V(H_0 + b) - V(H_0 + \bar{b})\right]$$  \hspace{1cm} (3.1')

where $\pi$ is the risk premium associated with the financial lottery:

$$\hat{\sigma}_{LF} \quad -\bar{C}$$

$$1 - \hat{\sigma}_{LF} \quad 0$$
Because the R.H.S. (right hand side) of (3.1') is necessarily non negative one has:

\[ \tilde{\sigma}_{LF} \tilde{C} + \pi \leq C \]

so that

\[ \frac{\tilde{\sigma}_{LF}}{C} \leq \frac{\pi}{C} < \frac{C}{C} \tag{3.2} \]

a result that will be used later on.

3.b. The first-best solution

In this environment the regulator determines who receives L.T.T. which is offered free of charge. Patients who do not have the L.T.T. and who become sick get the E.T. also free of charge. The program is financed by a tax T on each individual in order to cover the treatment costs. Formally one has:

\[ \max_{\tilde{\sigma}} U(W-T) + V(H_0 + b) \int_0^{\tilde{\sigma}} \sigma dF(\sigma) + V(H_0 + b) \int_0^{\tilde{\sigma}} \sigma dF(\sigma) + (1 - \tilde{\sigma}) V(H_2) \tag{3.3} \]

s.t. \[ T = C \int_0^{\tilde{\sigma}} \sigma dF(\sigma) d\sigma + C [1 - F(\tilde{\sigma})] \]

The first-order condition (F.O.C.) associated with this program is:

\[ (C - \tilde{\sigma}_{FB} \tilde{C}) U'(W-T) = \tilde{\sigma}_{FB} \left( V(H_0 + b) - V(H_0 + \tilde{b}) \right) \tag{3.4} \]

where \( \tilde{\sigma}_{FB} \) is the treatment threshold for the first-best solution.

The second order condition for a maximum is easily checked and notice again that because the R.H.S. of (3.4) is non negative, one also has a natural upper-bound for \( \tilde{\sigma}_{FB} \) which is:

\[ \tilde{\sigma}_{FB} \leq \frac{C}{C} \]

While (3.1') and (3.4) have rather similar structures it is pretty obvious that in general \( \tilde{\sigma}_{LF} \) and \( \tilde{\sigma}_{FB} \) cannot be compared, essentially because they are influenced by different and independent parameters. The threshold under L.F. \( \tilde{\sigma}_{LF} \) depends upon patients' risk aversion through \( \pi \) and is not at all dependent upon the distribution of \( \sigma \) in the population. On the contrary, under the first-best solution, \( \tilde{\sigma}_{FB} \) critically depends upon that distribution because T is linked to \( f(\sigma) \) and \( F(\sigma) \) through the budget constraint.

Although \( \tilde{\sigma}_{LF} \) and \( \tilde{\sigma}_{FB} \) cannot be compared in general, we now turn to a specific situation where the comparison can be made.
3.3. A specific case

When L.T.T. and E.T. have the same medical benefit \( \bar{b} = b \) an easy comparison between \( \hat{\sigma}_{LF} \) and \( \hat{\sigma}_{FB} \) emerges. In this case indeed the right hand side of \((3.1')\) becomes equal to zero so that:

\[
\hat{\sigma}_{LF} = \frac{C - \pi}{C} \tag{3.5}\]

while \((3.4)\) yields:

\[
\hat{\sigma}_{FB} = \frac{C}{C} \tag{3.5'}\]

As a result — and contrarily to a widespread belief — more patients adopt L.T.T. under the laissez-faire solution than under the first-best one. However, the reasons for the result can be easily understood. When \( b = \bar{b} \) the regulator who adopts the first-best solution is indifferent between the two treatments from a medical point of view and he tries to minimize the T value that permits every sick patient to reach a health level \( H_0 + b = H_0 + \bar{b} \). This minimization of T is obtained when \( \hat{\sigma}_{FB} = \frac{C}{C} \) i.e. when the marginal cost of obtaining the benefit \( b = \bar{b} \) for one sick patient is the same under each strategy\(^2\).

In the laissez-faire however, risk averse individual patients — at equal medical benefit — have a propensity to favor L.T.T. because it generates a known cost \( C \) as opposed to the random one induced by E.T. Hence they use L.T.T. more often than what would be observed in the first-best solution.

---

\(^2\) Indeed, to obtain \( b = \bar{b} \) for one sick patient with E.T., the marginal cost is \( \bar{C} \). However to obtain \( b = \bar{b} \) for one sick patient with L.T.T. this strategy has to be applied to \( \frac{1}{\sigma} \) patients inducing a cost equal to \( \frac{C}{\sigma} \).
SECTION 4: THE "SECOND-BEST" APPROACH AND ITS PROPERTIES

In a second best environment, treatment decisions are left to the patient but they are influenced by the regulator through his choice of \( \hat{s}, s \) and \( T \). The regulator selects these control variables in order to maximize the sum of individual utilities under a global budget restraint. Formally, one has:

\[
\begin{align*}
\text{Max}_{T,s,\hat{s},\lambda} \quad Z = & \int_{0}^{\hat{s}} \left\{ \sigma \left( U(W - T - (1 - \hat{s})\hat{C}) + V(H_0 + \hat{b}) \right) \\
& + (1 - \sigma) \left( U(W - T + V(H_2)) \right) dF(\sigma) \\
& + \int_{0}^{1} \left\{ U(W - T - (1 - s)C) + \sigma V(H_0 + b) + (1 - \sigma) V(H_2) \right\} dF(\sigma) \\
& + \lambda \left( T - \hat{s} \hat{C} \int_{0}^{\hat{s}} \sigma dF(\sigma) - sC(1 - F(\hat{s})) \right) \tag{4.1}
\end{align*}
\]

where \( \lambda \) is the Lagrange multiplier and where \( \hat{s} \) is defined as in (2.1).

The F.O.C. associated to the program are:

\[
\frac{\partial Z}{\partial T} = -E[U] + \lambda \left[ 1 - \hat{s} \hat{C} \frac{d\sigma}{dT} + sC \frac{d\sigma}{dT} \right] = 0 \tag{4.1'}
\]

\[
\frac{\partial Z}{\partial s} = C U' (W - T - (1 - s)C)(1 - F(\hat{s})) - \lambda \left[ C(1 - F(\hat{s})) + (\hat{s} \hat{C} - sC) f(\hat{s}) \frac{d\sigma}{ds} \right] = 0 \tag{4.1'\prime}
\]

\[
\frac{\partial Z}{\partial \hat{s}} = \hat{C} U' (W - T - (1 - \hat{s})\hat{C}) \int_{0}^{\hat{s}} \sigma dF(\sigma) - \lambda \left[ \int_{0}^{\hat{s}} \sigma dF(\hat{s}) + (\hat{s} \hat{C} - sC) f(\hat{s}) \frac{d\sigma}{d\hat{s}} \right] = 0 \tag{4.1''}
\]

\[
\frac{\partial Z}{\partial \lambda} = T - \hat{s} \hat{C} \int_{0}^{\hat{s}} \sigma dF(\sigma) - sC(1 - F(\hat{s})) = 0 \tag{4.1''\prime}
\]

These F.O.C. deserve a careful interpretation. Consider first (4.1'\prime). When \( s \) is increased coeteris paribus there is a gain in utility for all the patients who use the long term treatment. This number of patients is \( 1 - F(\hat{s}) \) and the gain in utility is measured by \( CU'(W - T - (1 - s)C) \). The marginal cost of raising \( s \) is represented by the other term in (4.1'\prime). Quite naturally it is made of \( C(1 - F(\hat{s})) \) which is the direct cost for the regulator of raising \( s \). However the increase in \( s \) has another (indirect) effect on the budget restraint which
is captured by \( (\hat{s} \hat{\sigma} \hat{C} - sC)f(\hat{\sigma}) \frac{d\hat{\sigma}}{ds} \). When \( s \) is increased more patients decide to get the long-term treatment. Their number is given by \( f(\hat{\sigma}) \) at the margin and they each cost \( sC \) to the regulator. Besides among these newcomers to the L.T. treatment a fraction \( \hat{\sigma} \) would have needed the emergency treatment if \( s \) had not been raised. Hence the increase in \( s \) generates an indirect saving for the regulator which amounts to \( \hat{s} \hat{\sigma} \hat{C} f(\hat{\sigma}) \).

Of course all these monetary consequences for the regulator are multiplied in (4.1’’) by \( \lambda \) which is linked to the marginal utility of wealth.

A similar interpretation applies to (4.1’’’). Notice, also, that in (4.1’) \( E[U'] \) is in general not equal to \( \lambda \) unless \( \frac{d\hat{\sigma}}{dT} = 0 \). Indeed when the regulator changes \( T \) this in general affects \( \hat{\sigma} \) with the implications for the regulator’s expenses that we have just described.

Although they do not lend themselves to an easy comparative statics exercise\(^3\), these F.O.C. already reveal some important features of the optimal co-payment rates. In fact it can be proven that it is never optimal to equate the co-payment rates and that systematically \( \hat{s}^* \) should exceed \( s^* \).

The proof is made in steps.

1. When \( s = \hat{s}, \hat{s} \hat{\sigma} \hat{C} - sC \) is negative. This result – which will be very important in step 2 – is easily obtained.

Indeed rewrite (2.1) by using the concept of the risk premium, i.e.:

\[
U(W - T - (1 - \hat{s})\hat{C} - \pi) - U(W - T - (1 - s)C) = \hat{\sigma} (V(H_0 + b) - V(H_0 + b)) \tag{4.2}
\]

Because the right hand side of (4.2) is non negative one has:

\[
(1 - \hat{s})\hat{C} - \pi < (1 - s)C
\]

which becomes:

\[
(1 - s)\hat{C} + \pi < (1 - s)C
\]

when \( s = \hat{s} \).

Dividing by \( (1 - s) \) on each side and using the fact that \( \pi \) is positive yields the result.

\(^3\) The S.O.C. for a maximum are not easily checked since they involve the slope of the density function \( f(\sigma) \) which can be of any sign.
(2) Consider now the regulator's optimality condition with respect to \( s \) and \( \bar{s} \)
((4.1")) for any pre-determined \( T \). After obvious manipulations they become:

\[
U'(W - T - (1 - s)C) = \frac{1 + \left( \bar{s} \bar{\sigma} \bar{C} - sC \right) \int_0^\infty \sigma dF(\sigma)}{C(1 - F(\bar{\sigma}))} \frac{d\bar{\sigma}}{ds} \\
U'(W - T - (1 - \bar{s})\bar{C}) = \frac{1 + \left( \bar{s} \bar{\sigma} \bar{C} - sC \right) \int_0^\infty \sigma dF(\sigma)}{\bar{C} \int_0^\infty \sigma dF(\sigma)} \frac{d\bar{\sigma}}{d\bar{s}}
\]

(4.3)

Now if \( s = \bar{s} \) observe:

- that the ratio on the left hand side of (4.3) is surely smaller than unity. Indeed when \( s = \bar{s} \) and because \( \bar{C} > C \):

\[ W - T - (1 - s)C > W - T - (1 - s)\bar{C} \]

Then because of risk aversion \( U'(W - T - (1 - s)C) \) is smaller than \( U'(W - T - (1 - s)\bar{C}) \);

- that the ratio on the right hand side of (4.3) is larger than unity. Indeed at \( s = \bar{s} \) we know that \( \bar{s} \bar{\sigma} \bar{C} - sC \) is negative. Besides using the results in (2.2) and (2.3) about the opposite signs of \( \frac{d\bar{\sigma}}{ds} \) and \( \frac{d\bar{\sigma}}{d\bar{s}} \), one easily finds that the numerator on the right hand side of (4.3) is larger than unity while the denominator is smaller than unity.

Since the L.H.S. of (4.3) is smaller than unity at \( s = \bar{s} \) while the reverse holds for the R.H.S. obviously \( s = \bar{s} \) cannot be an optimal solution for any predetermined \( T \) value.

This result is central for the paper since it clearly shows that the second best optimum cannot be reached when there is no differentiation in the co-payment rates.

(3) To satisfy (4.3) as an equality for any given \( T \), two strategies are possible:

- either increase \( s \) and decrease \( \bar{s} \);
- or decrease \( s \) and increase \( \bar{s} \).

If the first strategy is chosen it is obvious that an equality cannot be obtained. Indeed the increase in \( s \) lowers the numerator on the L.H.S. while the decrease in \( \bar{s} \) increases the denominator. Hence the L.H.S. of (4.3) remains surely smaller than unity. Besides from (2.2) and (2.3) the first strategy reduces \( \bar{\sigma} \) so that \( \bar{s} \bar{\sigma} \bar{C} - sC \) is more negative than before implying that the R.H.S. remains larger than unity.
Notice that the second strategy \( \{ \Delta s < 0 \text{ and } \Delta \bar{s} > 0 \} \) starting from \( s = \bar{s} \) induces a movement in the right direction to obtain the optimality condition in (4.3). Hence at any level of \( T \), including the optimal one, one necessarily has:

\[
s^* < \bar{s}^*
\]

which means that the co-payment rate for the emergency medicine should be smaller than for the long-term one.

This result may look surprising at first glance. Indeed because it is medically more efficient \( b \geq \bar{b} \) and less costly \( C \leq \bar{C} \) one might think that L.T.T. should be encouraged by a more generous subsidy rate. This is not so however essentially because the subsidy rates are a form of insurance for the patients. Since insurance is a very appropriate way to protect against catastrophic risks (small probability and large potential losses) and because emergency medicine has precisely these features (it is used by patients with low probabilities of sickners who may have to face a large cost (loss) if they do become sick) it is not surprising that it should be more generously covered by the public insurance scheme.
CONCLUSION

This paper is only a first step in the analysis of the optimal co-payment rate for different medical activities.

Quite obviously both medical strategies that were considered should be refined. One should also allow for an explicit consideration of other medical activities (such as prevention or diagnostic medicine).

Whatever the model adopted however, the case seems strong in favor of non-uniform co-payment rates.