Copenhagen
Business School
HANDELSHØJSKOLEN

Department of Economics

Copenhagen Business School

Working paper 5-2007

ACCOUNTING FOR MORAL HAZARD IN COST-EFFECTIVENESS ANALYSIS

Bodil Olai Hansen Hans Keiding

Accounting for Moral Hazard In Cost-Effectiveness Analysis

Bodil O. Hansen
Copenhagen Business School

Hans Keiding
University of Copenhagen

October 2007

Abstract

In the paper, we use the theory of mechanism design to exhibit the cost of efficient provision of healthcare, defined as the uniquely defined sum of individual side payments which would eliminate moral hazard. It is argued that this cost may be used to assess the costs arising from use of the treatment in cases where it is not appropriate from a strictly medical point of view. An example is given to indicate how this assessment might enter into practical cost-effectiveness analysis.

Keywords: cost-effectiveness, moral hazard, QALY, mechanism design.

JEL classification: I18, C72

1. Introduction

The increasing awareness of the cost of healthcare has led to the introduction of economic tools in decisions about treatments to be offered and reimbursement of patients' outlays. As a result of practical needs as well as theoretical considerations, a discipline of health economic assessments has developed, documented not only in numerous journal articles but also in textbooks such as Drummond et al. (1997), Drummond and McGuire (2001).

The basic tool for health economic assessment has always been the search for a cost-effectiveness ratio of a given treatment or medicine, and in recent years, the standard way of presenting results here is in the form of cost per QALY achieved. Much effort has been laid down in the development of methods for measuring QALY weights of health states, with the ultimate goal of presenting the decision maker with the best possible way of

utilizing the funds available for health care. The very terminology suggests that choosing the treatment with the lowest cost-effectiveness ratio implies that the maximal gain in health will be achieved; the decision is the most *effective* to be bought for the money available.

In the present paper we shall argue that even if we agree both on the objective of achieving effectiveness (maximum of health for the given amount of money) *and* accept the QALY approach as a valid measurement of health (a viewpoint which has went almost unchallenged for surprisingly many years, but where critical voices are now appearing), there can be no effectiveness without *efficiency,* in the sense that the interventions chosen should be applied to the right patients. Even if we accept the validity of QALY measurement, this efficiency aspect is *not* taken care of in a standard cost-effectiveness analysis.

What is missing in the standard approach is the behaviour of the patient and the prescribing doctor when the drug (or in general, the medical intervention) has been accepted as part of what is offered in the healthcare system. In this situation we may encounter overuse caused either by favorable reimbursement rates or simply by patient preferences for the new treatment, or underuse, again either by the provider's reluctance to use it or by patients' substitution towards cheaper though less effective drugs. This phenomenon is often referred to in practice as *diagnosis drift,* meaning that the drug is applied beyond its original designation, and decision makers are worried about it when deciding on marketing and reimbursement, but since it is not readily quantifiable and definitely not supported by any clinical data, it is rarely taken into account.

From the point of view of health economics, diagnosis drift is an instance of *moral hazard,* where individuals strive to achieve the best possible treatment given that their true state of health is only partially observable to the healthcare organization delivering the treatment. As is well-known, the phenomenon is not easily done away with, and the point of the present investigation is not to find ways of countering moral hazard, but rather to see whether there is a natural measure of the amount of moral hazard activity connected with a given treatment, which might be incorporated in the cost-effectiveness analysis of this treatment as a particular inefficiency cost.

We look at a somewhat simplified world, where QALY measurement is universally available and valid, so that even patients assess their health gain from any treatment in terms of QALY. This is clearly not a realistic assumption, but, given the almost unanimous acceptance of QALY measurement in the profession today, not an unusual one. Patients are not identical, and their treatment needs differ, but the healthcare system has as its objective to provide as much health gain to the population as possible, an objective which in our simplified world means that it aims at maximizing the total QALY gain of the population. A situation where total health needs are met in such a way is *efficient,* and without explicit

mentioning this seems to be the generally shared objective of contemporary healthcare systems.

Due to the presence of moral hazard, the healthcare system faces an information problem. Telling the truth is not necessarily in the interest of the patient (if some overdoing of the subjective symptoms can result in a preferred or cheaper treatment), and with biased information, the efficiency is endangered. In practice, what most concerns the decision makers seem to be the additional cost to the system which is caused by possible misrepresentation, but from an overall point of view the main problem is that naive methods of collecting information may fail.

This is a typical problem of *mechanism design*, and indeed our setup is such that we may apply some of the important results from the theory of mechanism design to our problem. We show that the information problem can indeed be done away with if we are willing to include side payments, either from the healthcare organization or from the patients.

2. The model; assessing the cost of efficiency in treatment

We consider a community with N individuals, interpreted as potential patients. We assume that patients' health states belong to a set H, the structure of which will be specified later, except for the detail that it contains a distinguished element denoted h_0 and interpreted as "full health". A health profile is an array $\boldsymbol{h} = (h^1, \ldots, h^N)$ specifying the health state of each individual. The community has a healthcare organization which may provide treatments $t \in T$ to individuals; we assume that also T contains a distinguished element 0 interpreted as "no treatment". A treatment plan is an element of T^N, where for $\boldsymbol{t} = (t^1, \ldots, t^N) \in T^N$, the ith coordinate specifies the treatment of individual i. The gain in terms of health-related quality of life obtained by individual i in health state h obtaining treatment t is denoted $Q^i(t, h)$. We assume that $Q^i(t, h_0) = 0$ for any treatment t: treatment has no effect on patients in perfect health. An example of a utility indicator of this form is the QALY gain from treatment t at health state h.

Assume that the healthcare provider has limited capacity, so that the treatment plan must belong to a given subset B of T^N. In the traditional approach to cost-effectiveness analysis, the set B is all the alternative possibilities of treating the population for different illnesses that can be financed with the budget which is available to the decision maker who represents the healthcare system. If treatments are provided under the criterion of bringing maximal benefit to as many individuals as possible, an optimal treatment plan $\hat{\boldsymbol{t}}(\boldsymbol{h})$ given the type profile $\boldsymbol{h}$ is one which gives highest aggregate utility to the population,

that is satisfies

$$\hat{\boldsymbol{t}}(\boldsymbol{h}) \in \text{argmax}_{(t^1,\ldots,t^N)\in B} \sum_{i=1}^{N} Q(t^i, h^i). \tag{1}$$

In the case where QALY measures give a faithful picture of health-related quality of life, so that each individual has utility function $Q(t, h)$, the above method of choosing treatment plan is the utilitarian decision rule (society's choice maximizes the sum of the individual utilities).

As argued in the preceding section, systems of healthcare provision are susceptible to moral hazard, in the sense that patients might want a treatment which increases personal satisfaction but does not represent the best possible way of utilizing limited resources. In the context of our model, it means that individuals will select their messages concerning health state in such a way that the resulting treatment plan is as good as possible for themselves. This has the unfortunate side-effect of distorting the decision which no longer necessarily represents the optimal treatment plan given the true health profile $\boldsymbol{h}$.

Since moral hazard in our present setup takes the form of misrepresentation of types h^i by patients, it seems reasonable to consider methods for obtaining truthful revelation. The most obvious candidate here is the Vickrey-Clarke-Groves (VCG) mechanism, in which the statements of type give rise not only to an allocation (namely the treatment plan given by (1)) but also to a system of (user) payments from each of the individuals. For the VCG mechanism to work properly, we shall assume that the utility of each individual, depending on type (health), treatment, and money outlays, is quasi-linear, taking the form

$$u(t^i, h^i, p^i) = Q(t^i, h^i) - p^i.$$

The VCG mechanism then consists of the allocation part as specified in (1), using the type statements (not necessarily truthful) of the individuals, and the payment part, whereby individual i pays

$$p^i(\boldsymbol{h}) = \sum_{j\neq i} Q(\hat{\boldsymbol{t}}^j(h_0, h^{-i}), h^j) - \sum_{j\neq i} Q(\hat{\boldsymbol{t}}^j(\boldsymbol{h}), h^j) + g^i(h^{-i}), \tag{2}$$

g^i being an arbitrary function depending only on the types stated by the individuals different from i. Since the functions g^i are not otherwise specified, there is a whole family of VCG mechanisms available.

We may now exploit the classical following results from the theory of mechanism design:

PROPOSITION 1. *(a) In the mechanism specified by (1) and (2), it is a weakly dominant strategy for each individual i to state the true health state.*

(b) If the set H of health states is a convex subset of $\mathbb{R}^d$ for some d, then any mechanism such that truth is a weakly dominant strategy for each individual is a VCG mechanism.

Part (a) of Proposition 1 was established by Clarke (1971), Groves (1973), and part (b) is due to Green and Laffont (1977) and Holmström (1979). Proof of the proposition can be found in many texts, e.g. in Milgrom (2004), so we omit it here.

It may be instructive to consider the payment formula somewhat closer, choosing the simplest version with the functions g^i being identically 0, the so-called pivot VCG mechanism. In this case, what is demanded from i as user payment is the difference between the aggregate health gain to be obtained by the others if individual i has no need of treatment whatsoever, and the health gains obtainable for the others if also individual i must be treated. Using a terminology which is well-known among health economists, we have that i must cover the *opportunity cost* of her own treatment, measured in health gains. It should be noticed that this opportunity cost, measured as the health gain that all other patients lose when a given patient is treated, will typically be smaller than the full cost of treatment.

It seems reasonable to measure the cost of efficiency, at least in the context of a particular chosen mechanism for attaining it, as the sum of the payments by individuals. We shall use the terminology *society's efficiency cost* for this measure. Assuming that patients reporting the state h_0 corresponding to full health and who will not change the treatment profile $\boldsymbol{t}$ by their presence, will pay 0, this efficiency cost (considered on a per potential patient basis) will be

$$c_{soc}(h^i, h^{-i}) = \sum_{j \neq i} Q(\hat{\boldsymbol{t}}^j(h_0, h^{-i}), h^j) - \sum_{j \neq i} Q(\hat{\boldsymbol{t}}^j(\boldsymbol{h}), h^j).$$

Taking expectation over the distribution of health states of the remaining individuals, given the health state h^i of individual i, we get the expected payment of i in health state h^i,

$$c_{soc}(h^i) = \mathsf{E}\left[c_{soc}(h^i, \tilde{h}^{-i})\right]; \tag{3}$$

Finally, taking expectation over health states of patients with $h^i \neq h_0$ we get

$$C_{soc} = \mathsf{E}\left[c_{soc}(h^i) | h^i \neq h_0\right]$$

as an assessment of the per patient cost of achieving efficient allocation of treatments. It should be emphasized that this cost is not related to treatment of patients (as the direct costs) nor to patients' functioning in production or consumption (as indirect cost). They are purely theoretical in the sense that these costs *would have been incurred* if patient treatment was to be efficient, which in practice it is not. Thus, these are *virtual costs* of misallocation, the cost that would be incurred if misallocation was to be set right. We shall return in a later section to a discussion of this somewhat exotic type of cost and their possible role in cost-effectiveness analysis.

What may rightly be objected at this point is that in the approach to efficient allocation, we have insisted on a particular mechanism, or rather, in view of Proposition 1(b), on a particular solution concept, namely equilibrium in (weakly) dominant strategies, meaning that truthful reporting is as good as any other report no matter what the other individuals report, and better in some cases. We might settle down with a mechanism which achieves efficient allocation of treatment using a less demanding equilibrium concept, and this might change our assessment of the efficiency cost. We shall see in the next section that this is not the case.

3. Independence of the assessment

The reliance on a particular mechanism, namely the VCG mechanism, for achieving efficient treatment, might seem a weakness of the approach, so we should consider possible alternatives. Also, the particular property of the VCG mechanism that truthtelling is weakly dominant may conceivably lead to higher than necessary user payments. We therefore consider an arbitrary mechanisms $\mathcal{M} = (S, \tau, \pi)$, where S is an abstract space of messages, $\tau : S^N \to T^N$ an outcome map which assigns a treatment plan to each array of messages, and $\pi : S^N \to \mathbb{R}^N$ a payment function specifying the payment for each individual given the messages sent by all individuals.

A *Bayesian equilibrium* is a family of strategies, which are health-dependent messages $(\sigma^i)_{i\in N}$, where each $\sigma^i : H \to S$ assigns a message to each health state, such that for each i and h^i, $\sigma^i(h^i)$ maximizes

$$U(h^i; (\sigma^i)_{i\in N}) = \mathsf{E}\,[Q(\tau(\sigma^i(h^i), (\sigma^j(\tilde{h}^j))_{j\neq i}), h^i) - \pi_i(\sigma^i(h^i), (\sigma^j(\tilde{h}^j))_{j\neq i})],$$

the expectation being taken with respect to the conditional distribution of h^{-i} given h^i, over all messages $s \in S$. Thus, in a Bayesian equilibrium, each individual chooses her message in such a way that average gain, taking both allocation and payment into account, is as great as possible; here the average is taken over all the health configurations of the remaining individuals, using the prior knowledge of its overall distribution. Clearly, a strategy may be best against average behaviour of the others without being best against any single configuration of the remaining individuals. On the other hand, if a strategy is weakly dominant, being best in any configuration, then it is also best on the average and therefore a Bayesian equilibrium.

For the following result, we need to specify the nature of the set H of possible health states of each individual, which we assume to be the unit interval $[0, 1]$. Thus, we assume that health can be assessed on a one-dimensional scale; however, we need not assume any properties of this scale for our results, so that our assumption is much less restrictive that

assuming e.g. QALY measurement of health states (as distinct from QALY measurement of health improvement from treatment in any given health state, which we have assumed above).

PROPOSITION 2. *Let $\mathcal{M}$ be a mechanism, and let σ be a Bayesian equilibrium such that $\tau(\sigma(\boldsymbol{h})) \in \hat{\boldsymbol{t}}(\boldsymbol{h})$ for each $\boldsymbol{h} \in H^N$. Assume that $H = [0,1]$ with $h_0 = 0$, and that the expected payment of a healthy individual $\mathsf{E}\left[\pi(\sigma^i(h_0), (\sigma^j(\tilde{h}^j))_{j\neq i})\right]$ is 0. Then*

$$\mathsf{E}\left[\pi(\sigma^i(h^i), (\sigma^j(\tilde{h}^j))_{j\neq i})\right] = c_{soc}(h^i)$$

for each i, where $c_{soc}(h^i)$ is defined in (3).

PROOF: Let i be arbitrary, and assume that the profile of health states $\boldsymbol{h}$ is given. Using Myerson's lemma (see e.g. Milgrom (2004), p.74) and writing

$$\begin{aligned}
\pi^*(\boldsymbol{h}) &= \pi(\sigma^i(h^i), (\sigma^j(\tilde{h}^j))_{j\neq i}),\\
\tau^*(\boldsymbol{h}) &= \tau(\sigma^i(h^i), (\sigma^j(\tilde{h}^j))_{j\neq i}),\\
U^*(h_0) &= \max_{\hat{\sigma}^i} U(h_0; (\hat{\sigma}^i, (\sigma^j)_{j\neq i})),
\end{aligned}$$

we have that the expected payment of individual i has the form

$$\begin{aligned}
\mathsf{E}\left[\pi^*(\tilde{\boldsymbol{h}})|\tilde{h}^i = h^i\right] = & -U^*(h_0) + \mathsf{E}\left[Q(\tau^*(\tilde{\boldsymbol{h}})|\tilde{h}^i = h^i\right]\\
& - \int_{h_0}^{h^i} \left(\sum_{t^i \in T} \text{Prob}\{\tau^*(\tilde{\boldsymbol{h}}) = t^i | \tilde{h}^i = s\} Q_2'(t^i, s) \right) ds.
\end{aligned}$$

For $h^i = h_0$, the expression reduces to

$$\mathsf{E}\left[\pi^*(\tilde{\boldsymbol{h}})|\tilde{h}^i = h_0\right] = -U^*(h_0) + \mathsf{E}\left[Q(\tau^*(\tilde{\boldsymbol{h}})|\tilde{h}^i = h_0\right].$$

Here, the term $U^*(h_0)$ depends on the mechanism, but the second term on the right-hand side depends only on the allocation rule, and since $\tau^*(\boldsymbol{h}) \in \hat{\boldsymbol{t}}(\boldsymbol{h})$ for all $\boldsymbol{h}$, we have that $\mathsf{E}\left[Q(\tau^*(\tilde{\boldsymbol{h}})|h^i = h_0\right] = 0$. We then obtain that

$$U^*(h_0) = -\mathsf{E}\left[\pi^*(\tilde{\boldsymbol{h}})|\tilde{h}^i = h_0\right] = 0$$

since expected payment in health state h_0 is 0 by our assumptions. It follows now that expected payment in any health state depends only on the allocation rule, and since this is the same as in the VCG mechanism, we get the result. □

4. A simplified assessment formula

In the previous sections, we have developed a theoretical measure of the cost of less-than-optimal use of the system for healthcare provision. It remains to connect this measure to information which is available when carrying out the cost-effectiveness analysis, which in most cases means before the drug has went into use, so that no empirical data on prescription practice is available. Also, practical usefulness implies that the information needed should relate to the treatment considered rather than, as has been the case above, to the healthcare system as a whole and the totality of its treatments.

As we shall show in this section, if we are willing to accept some further simplifying assumptions, then there is a rather simple way of assessing the moral hazard cost of a treatment. First of all, we shall make an assumption of separability of treatments:

ASSUMPTION A1. *There is a family A of interventions such that*

(i) $T = \prod_{\alpha \in A} T_\alpha$, *and*

(ii) $Q(t,h) = \sum_{\alpha \in A} Q(t_\alpha, h)$, *each* $h \in H$ *and* $t = (t_\alpha)_{\alpha \in A}$,

(iii) $B = \cap_{\alpha \in A} B_\alpha$ *for given subsets* B_α *of* $(T_\alpha)^N$.

What is assumed here is that treatments which can be given to patients are fully described by a (finite) menu of interventions; this first part of the assumption is therefore not particularly restrictive. The second part of our assumption states that the total utility gain of a treatment is the simple sum of the utility gains of the interventions of which the treatment consists; this assumption is implicit in the traditional QALY approach to effect measurement.

For ease of notation, we identify the treatment t_α in the intervention α with the treatment $t = (0, \ldots, 0, t_\alpha, 0, \ldots, 0)$ writing $Q(t_\alpha, h)$ for the utility gain obtained in health state h by receiving only treatment t_α from intervention α and nothing else.

The final part of Assumption 1 states that the set of feasible treatments can be seen as consisting of feasible sets pertaining to each intervention. Once again this is a separability assumption which is useful for assessment purposes but does not quite correspond to reality, in particular if B is interpreted as a set of treatments which can be carried through with a given budget. Here we reason as if the budget had already been split into separate budgets for each intervention, which may work if the treatment plan is already known to be overall welfare maximizing, but not in general.

The main advantage of our last assumptions is that they allow us to consider the efficiency problem for each intervention separately rather than for the system of healthcare provision as a whole. This property is formulated below as a lemma.

LEMMA 1. *Under A1, a treatment plan* $(\hat{t}^1, \ldots, \hat{t}^N)$ *is optimal given the profile of health states* $\boldsymbol{h}$ *if and only if for each* $\alpha \in A$, *the intervention profile* $(\hat{t}^1_\alpha, \ldots, \hat{t}^N_\alpha)$ *maximizes*

$\sum_{i=1}^N Q(t_\alpha^i, h^i)$ *on* B_α.

PROOF: Let $\boldsymbol{h}$ be arbitrary. It is trivial that if $\hat{\boldsymbol{t}}$ with $\hat{t}^i = (\hat{t}_\alpha^i)_{\alpha \in A}$ for each i satisfies (1), then $\hat{\boldsymbol{t}}_\alpha = (\hat{t}_\alpha^1, \ldots, \hat{t}_\alpha^N)$ satisfies

$$\hat{\boldsymbol{t}}_\alpha \in \text{argmax}_{(t_\alpha^1, \ldots, t_\alpha^N) \in B_\alpha} \sum_{i=1}^N Q(t^i, h^i). \tag{4}$$

Suppose conversely that (4) is satisfied for each $\alpha \in A$. Then for any treatment plan $\boldsymbol{t}$ we have $\sum_{i=1}^N Q(t_\alpha^i, h^i) \leq \sum_{i=1}^N Q(\hat{t}_\alpha^i, h^i)$ for all α, and by (ii) of Assumption 1, we get that

$$\sum_{i=1}^N Q(t^i, h^i) = \sum_{i=1}^N \sum_{\alpha \in A} Q(t_\alpha^i, h^i) \leq \sum_{i=1}^N \sum_{\alpha \in A} Q(\hat{t}_\alpha^i, h^i) = \sum_{i=1}^N Q(\hat{t}^i, h^i),$$

showing that $\hat{\boldsymbol{t}}$ satisfies (1). □

Using the result of Lemma 1, we may proceed towards the assessment of the cost of moral hazard concentrating upon the intervention and neglecting repercussions on the health system in general. Even so, the informational requirements of a cost computation are beyond what can be met in practice, and we shall simplify further, assuming that the treatment possibilities with the intervention considered reduce to a question of whether or not to treat, and that there are well-defined health levels above which this treatment are worthwhile for the patient and for the system, respectively.

PROPOSITION 3. *Assume A1, let individual health h be elements of $\mathbb{R}$, and let $\alpha \in A$ be an intervention with $T_\alpha = \{0, 1\}$. Assume that $Q(t, \cdot)$ is nonincreasing for each $t \in T_\alpha$ and that the health states of the individuals are independent with distribution function F admitting a continuous density, and that the population is large (N going to ∞). Finally, assume that the set B_α of feasible treatments is symmetric in N, so that $(t^1, \ldots, t^N) \in B_\alpha$ implies that $(t^{\sigma(1)}, \ldots, t^{\sigma(N)}) \in B_\alpha$ for each permutation σ of N.*

Then for each profile of health states $\boldsymbol{h}$ there is $\boldsymbol{t} \in \hat{\boldsymbol{t}}(\boldsymbol{h})$ and a threshold $\bar{h}(\boldsymbol{h})$ such that

$$t_\alpha^i = 1 \text{ if and only if } h^i \geq \bar{h}(\boldsymbol{h}), \tag{5}$$

and

$$C_{soc} = \frac{F(\bar{h}(\boldsymbol{h}))}{F(h^*)} Q(1, \bar{h}(\boldsymbol{h}), \tag{6}$$

where $h^* = \inf\{h \mid Q(1, h) > 0\}$ is the lower limit of the health states for which treatment matters.

The statement of the proposition looks somewhat abstract, but the result is easily interpreted: For each given profile of health states there is a threshold such that in the optimal treatment profile, individuals should be given treatment if and only if their health falls below this threshold (recall that for technical reasons we measure health as nonnegative real numbers with 0 as perfect health). Moreover, the social loss due to moral hazard can be evaluated as expected gain of treatment to an individual who is just below the treatment threshold.

PROOF OF PROPOSITION 3: Let the profile of health states $\boldsymbol{h}$ be given, and let $\hat{\boldsymbol{t}}(\boldsymbol{h})$ be the set of optimal treatment profiles at $\boldsymbol{h}$. By Lemma 1, any intervention profile $(\hat{t}_\alpha^1, \ldots, \hat{t}_\alpha^N)$ with $\hat{t} \in \hat{\boldsymbol{t}}(\boldsymbol{h})$ maximizes $\sum_{i=1}^N Q(t_\alpha^i, h^i)$ on B_α. Assume now that there are individuals i_1 and i_2 with $h^{i_1} < h^{i_2}$ but such that $\hat{t}_\alpha^{i_1} = 1$, $\hat{t}_\alpha^{i_2} = 0$. Then the treatment profile obtained from $(\hat{t}_\alpha^1, \ldots, \hat{t}_\alpha^N)$ by permuting i_1 and i_2 is feasible by symmetry of B_α, and since $Q(t, \cdot)$ is nonincreasing, we may assume that the optimal intervention profile has $\hat{t}_\alpha^{i_1} = 0$, $\hat{t}_\alpha^{i_2} = 1$. Consequently, the set of health states h such that $\hat{t}_\alpha^i = 1$ for some i with $h^i = h$ may be taken to be an interval $[h^*, \infty)$. This proves (5) and thereby the first part of the proposition.

For the second part, we begin by evaluating the individual payment in the VCG mechanism of individual i in health state h^i, given that the true health state has been reported by everyone. If $\hat{t}_\alpha^i = 1$, then withdrawing individual i will mean that a patient with health state just above the threshold $\bar{h}$ will get treatment instead of i, resulting in a utility gain of $Q(1, \bar{h})$, which would be the payment extracted from i in the VCG mechanism. If individual i is not treated, then eliminating i has no effect on the utility of the others, and the payment is zero.

By our assumption of large population and independence of health distribution, we may assume that the treatment threshold is non-random, and since the payment $c_{soc}(h^i, h^{-i})$ depends on h^{-i} only through $\bar{h}$, we obtain C_{soc} by taking expectations over h^i, giving the desired formula (6). □

The simplification obtained in Proposition 3 is substantial. Instead of assessing payment formula in abstract mechanisms, we need only look for

(i) the treatment threshold,

(ii) the proportion of relevant patients whose health state is below this threshold, so that they are entitled to treatment, and

(iii) the utility gain from treatment at the threshold.

These quantities can usually be assessed easily from available information, at least as approximations. Clearly, the threshold as defined above presupposes a maximization of total utility gains from treatment, whereas observable thresholds are determined by clinical considerations; approximating theoretical concepts by real-life quantities determined in

a different way is however a practice which is widespread in cost-effectiveness analysis. The remaining quantities are in most cases found already in standard practice and therefore easily accessible.

As can be seen from (6), the cost of moral hazard, as expressed by C_{soc}, is a product of two factors, the first being the relative frequency of patients susceptible to moral hazard which are actually treated, and the other one being the utility gain from treatment at the limit. It should be noticed that by our assumption of quasi-linearity, this utility gain is measured in money units, so that $Q(1, h)$ represents the money equivalent of the QALY gain which an individual in health state h derives from treatment. There is no standard way of assigning money equivalents to QALY gains, but a commonly used method takes as its point of departure the rules-of-thumb guiding the healthcare organizations when adopting new interventions and treatments, according to which ¤ 30,000 is the limiting value of obtaining a QALY; identifying marginal valuation by price, this can be taken as an approximation of the money equivalent of a QALY.

For the evaluation of the first factor in (6), the frequency of treated individuals relative to all individuals who may indulge in moral hazard, this frequency should reflect the importance of moral hazard for the intervention considered. If, for example, the extent to which patients may misrepresent true state of health is restricted to an interval $\triangle$, then the quantity in the numerator reduces to $F(\underline{h} + \triangle) - F(\underline{h})$, giving a correspondingly lower social cost of moral hazard.

5. An application: General versus individual reimbursement for drugs

In the preceding sections, we have developed a method for assessing the cost of preventing moral hazard by voluntary means. The cost as found above expresses the total payments to be extracted from the relevant patients (treated or not) in the cheapest possible arrangement by which the welfare optimum is achieved when individuals choose what is best for themselves given the rules of this arrangement. The question arises as to whether this purely theoretical construction can shed light on practical matters; the purpose of this section is to argue that indeed it can.

The obvious way in which to use the look at the *relative* importance of moral hazard in any given case is to compare the cost of moral hazard as found above with the surplus cost of treating patients which should not have been treated in the welfare optimum. If the latter is small compared to the first one, then moral hazard is a minor problem – doing away with it would cost much more than accepting the welfare loss caused by it – whereas moral hazard becomes increasingly burdensome when the cost of over-treatment gets larger compared to the cost of preventing moral hazard.

Taking advantage of the results obtained in the previous section, we can get an approximate numerical assessment of these quantities in cases where the available data shed light on the distribution of benefits from treatment in the patient population.

PROPOSITION 4. *Let the cost δ of an intervention be given, and let G be the probability distribution of the cost-effectiveness ratio of the intervention. If q^* denotes the threshold for cost-effectiveness ratios of socially acceptable interventions, then the cost of over-treating patients is*

$$M_O = \delta(1 - G(q^*)),$$

whereas the cost of preventing moral hazard is

$$M_P = G(q^*)\delta\frac{1}{q^*},$$

both measured per capita of potential patients, so that the relative importance of moral hazard is

$$\frac{M_O}{M_P} = \frac{1 - G(q^*)}{G(q^*)}q^* \qquad (5).$$

PROOF: The expression for M_O is a simple consequence of the definition, since $1-G(q^*)$ is the fraction of patients for which treatment is socially suboptimal, but which are nevertheless treated due to moral hazard. For M_P, we use properties of the VCG mechanism: In optimum, the payment of truth-telling patients who are not treated is zero, since optimum remains unchanged if they are removed from the decision making; for patients treated, the payment is the difference between total satisfaction of the others when this patient is absent, and the actual total satisfaction of the others, and since removal of a patient treated means that a previous untreated patient gets treatment, the satisfaction of this patient must be at the threshold level of the cost-effectiveness ratio, meaning that payment of a treated patient is δ/q^*. This gives us the expression for M_P as stated in the proposition. □

It should be noticed that the quantity q^* which occurs in the expressions of Proposition 4 is a constant which does not depend on the intervention considered. Therefore, we may use the first factor on the right hand side in (5) as a measure of relative importance of moral hazard,

$$\rho_M = \frac{1 - G(q^*)}{G(q^*)}.$$

Comparing the values of ρ_M among different interventions we may then identify those for which moral hazard is a problem (high values of ρ_M) and consider whether some additional arrangements, not necessarily voluntary, may be in order to reduce the impact of moral hazard.

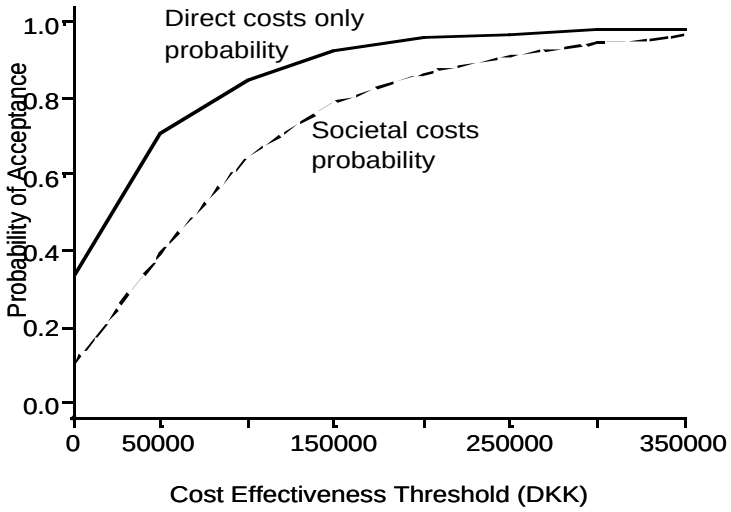

Figure 1

It so happens that the probability distribution G used in Proposition 4 is often available, at least approximately, as a result of the cost-effectiveness analysis of the intervention, namely in the form of the so-called cost-effectivity *acceptability curve* introduced by van Hout et al. (1994). Below we present an example of such a case.

The example considered is based on the work by Keiding et al. (2006), performing a cost-effectiveness analysis of losartan versus atenolol for treatment of patients with hypertension, based on patient data collected in the LIFE study (see Dahlöf et al., 2002). The basic results of the analysis was a cost-effectiveness ratio of DKK 19,668 per QALY when changing from atenolol to losartan if only direct cost is taken into consideration (analysis performed from the point of view of the healthcare organization) and DKK 72,564 per QALY gained if all costs are included (society's point of view).

Since the LIFE study included a very large number of patients, it was possible to carry out extensive statistical analysis, including in particular construction of cost-effectiveness acceptability curves, shown in Figure 1. Taking the informal rule-of-thumb for acceptable interventions, namely DKK 300,000 per QALY gained, we have that the probability of a patient with an effect of treatment which is below this threshold is 0.95 (all costs taken into account). From this we get a measure of relative importance of moral hazard of

$$\rho_M = \frac{0.05}{0.95} = 0.0526.$$

Roughly speaking, disregarding a constant which does not depend on the intervention

considered, the cost to society of accepting overconsumption is only about 5% of the cost incurred if the socially optimal level of treatment was to be achieved by voluntary decisions. In other words, accepting the moral hazard seems to be rather unimportant in the case considered.

It should be mentioned that the computations are based on several simplifications, both regarding the model (quasi-linear utilities, independence of treatments, selfish preferences etc.) and the approximations used. In particular, the interpretation of the cost-effectiveness acceptability curve as a probability density function of individual cost-effectiveness ratios is not quite exact, cf. Fenwick, O'Brien and Briggs (2004). However, even with its inherent shortcomings the method outlined presents a possibility for assessing the importance of moral hazard in each individual case, supplementing the general impressions which hitherto has been the basis for decision making.

6. Conclusion

In the present paper, we have developed a method for measuring the impact of moral hazard in situations where a given intervention may be demanded by patients who are not considered to benefit from this intervention to a sufficient extent to justify the cost of treatment to society. Since the phenomenon to be measured has the form of a deviation from an abstract (and typically not realized) social optimum, the measurement must necessarily take a somewhat roundabout approach.

The approach chosen exploits some basic properties of mechanism design, starting with the classical Vickrey-Clarke-Groves mechanism, which, if implemented, would result in socially optimal decisions (under the – unfortunately rather restrictive – assumptions of quasi-linear utility). The cost to the individuals of implementing the mechanism can be seen as a cost of doing away with moral hazard, and this interpretation is reinforced by the classical result from mechanism theory stating that any mechanism which achieves the social optimum in Bayesian equilibrium will give the same expected cost to the participants. With this notion of expected cost of moral hazard we can now assess the burden to society of this phenomenon; it turns out that in cases usually treated in cost-effectiveness studies, the equilibrium payments of the VCG mechanism can be assessed, at least up to a multiplicative constant, using data which are often available from the clinical trial. This will make it possible to compare interventions with regard to their vulnerability from moral hazard.

Needless to say, the approach is quite crude and many refinements are possible. Also, the method should be considered from an axiomatic point of view, listing desirable properties of any measure of the extent of moral hazard and characterizing this particular one by some such properties. This will be a matter of future research.

References

Dahlöf, B., R.B.Devereux, S.E.Kjeldsen et al. (2002), Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE): a randomized trial against atenolol, Lancet 359, 995 – 1003.

Drummond,M., B.J.O'Brien, G.L.Stoddart, G.W.Torrance (1997), Methods for the economic evaluation of health care programs, 2nd ed., Oxford University Press, Oxford.

Drummond,M. and A.McGuire (2001), Economic Evaluation in Health Care, Merging theory with practice, Oxford University Press, Oxford.

Clarke, E.H. (1971), Multipart pricing of public goods, Public Choice 11, 17 – 33.

Fenwick, E., B.J.O'Brien and A.Briggs (2004), Cost-effectiveness acceptability curves .. facts, fallacies and frequently asked questions, Health Economics 13, 405 – 415.

Green, J. and J.-J. Laffont (1977), Characterization of satisfactory mechanisms for the revelation of preferences for public goods, Econometrica 45, 427 – 438.

Groves, T. (1973), Incentives in teams, Econometrica 41, 617 – 631.

Holmström, B. (1979), Groves schemes on restricted domains, Econometrica 47, 1137 – 1144.

Keiding, H., P.Hildebrandt, T.Burke and G.W.Carides (2006), Cost-effectiveness of treatment for hypertension with losartan in Denmark (in Danish), Ugeskrift for Læger 168(42), 3623 – 3626.

Milgrom, P. (2004), Putting auction theory to work, Cambridge University Press, Cambridge.

Myerson, R.B. (1981), Optimal auction design, Mathematics of Operations Research 6, 58 – 73.

Van Hout, B.A., M.J.Al, G.S.Gordon and F.F.H.Rutten (1994), Costs, effects and c/e-ratios alongside a clinical trial, Health Economics 3, 309 – 319.