

FUTURE COSTS, FIXED HEALTHCARE BUDGETS, AND THE DECISION RULES OF COST-EFFECTIVENESS ANALYSIS

PIETER VAN BAAL^{a,*}, DAVID MELTZER^b and WERNER BROUWER^a

^a*Institute of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands*

^b*University of Chicago, Chicago, IL, USA*

ABSTRACT

Life-saving medical technologies result in additional demand for health care due to increased life expectancy. However, most economic evaluations do not include all medical costs that may result from this additional demand in health care and include only future costs of related illnesses. Although there has been much debate regarding the question to which extent future costs should be included from a societal perspective, the appropriate role of future medical costs in the widely adopted but more narrow healthcare perspective has been neglected. Using a theoretical model, we demonstrate that optimal decision rules for cost-effectiveness analyses assuming fixed healthcare budgets dictate that future costs of both related and unrelated medical care should be included. Practical relevance of including the costs of future unrelated medical care is illustrated using the example of transcatheter aortic valve implantation. Our findings suggest that guidelines should prescribe inclusion of these costs. Copyright © 2014 John Wiley & Sons, Ltd.

Received 24 March 2014; Revised 25 October 2014; Accepted 20 November 2014

KEY WORDS: future costs; economic evaluation; decision rules

1. INTRODUCTION

Medical interventions that increase life expectancy of patients may result in additional consumption of medical goods and services in so-called added life years. Added years are those years that would not have been lived without the intervention. Whereas part of this medical consumption in added years is directly related to the intervention, other costs in added years are not. The latter costs are normally labeled as costs of ‘unrelated medical care’.¹ An example of such costs would be those of treating dementia in added years due to successful cancer treatment. Although there is broad agreement that future related medical costs should be included in economic evaluations of healthcare interventions, there has been some debate in the theoretical literature regarding the inclusion of future unrelated medical costs in economic evaluations (Rappange *et al.*, 2008). Most of this discussion used a standard economic framework to answer the question to what extent future costs need to be included in evaluations in order to ensure social welfare maximization (Garber and Phelps, 1997; Meltzer, 1997; Feenstra *et al.*, 2008; Lee, 2008; Meltzer, 2008). Using representative consumer models, it has been shown that including all future costs (both related and unrelated medical costs as well as future non-medical costs) is required in order to maximize lifetime expected utility (Meltzer, 1997). Only under strong assumptions can a case be made to exclude future unrelated medical costs (Garber and

*Correspondence to: Institute of Health Policy and Management, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: vanbaal@bmg.eur.nl

¹Note that in the literature, these costs are also called indirect medical costs. Here, we will use the terms future unrelated medical costs/unrelated medical costs/indirect medical costs interchangeably.

Phelps, 1997; Meltzer, 1997; Feenstra *et al.*, 2008; Lee, 2008; Meltzer, 2008). However, to date, no attempts have been made to rigorously analyze the role of future medical costs specifically within a healthcare perspective in which healthcare budgets are exogenously set on a periodic basis. This is rather surprising given that in many jurisdictions, economic evaluations are carried out from a healthcare perspective, most prominently in England and Wales (Tarn and Smith, 2004). Hence, this paper aims to provide such an analysis.

It is important to note that although there has been much discussion from a theoretical perspective on the role of future medical costs, economic evaluations in practice often model the lifetime healthcare costs of related diseases but exclude the costs of unrelated diseases. In the most recent quantitative review (Meltzer, 2006), only about 50 cost-effectiveness studies had included future costs out of perhaps 1000 cost-effectiveness articles published in the prior decade. In some cases, this practice is the result of official guidelines for cost-effectiveness analysis. In several jurisdictions, it is explicitly advised to exclude the costs of unrelated medical care (Tarn and Smith, 2004). This is for instance the case in England and Wales, where the National Institute for Health and Care Excellence (NICE, 2013) guidelines apply. Although NICE guideline 5.5.1 suggests that all relevant costs (whether it be related or unrelated) for the National Health Service (NHS) should be included, 'For the reference case, costs should relate to resources that are under the control of the National Health Service (NHS) and Personal and Social Services (PSS)', there is a special clause dealing with costs in added years in guideline 5.5.7: 'Costs related to the condition of interest and incurred in additional years of life gained as a result of treatment should be included in the reference-case analysis. Costs that are considered to be unrelated to the condition or technology of interest should be excluded.' Justification for having guidelines recommending the exclusion of future unrelated medical costs, even those falling on future NHS budgets, is lacking. Therefore, we can only speculate that these guidelines are based on more informal arguments to exclude future unrelated medical costs such as mentioned in health economics textbooks (Drummond *et al.*, 2005; Morris *et al.*, 2007). For instance Morris *et al.* (2007) note that 'a cogent argument against the inclusion of future costs is that these relate to specific programmes or treatments which in themselves ought properly to be the subject of economic evaluation'. Other arguments mentioned in these textbooks are that it is difficult to estimate these costs and that their size is negligible. It is clear that such arguments deserve theoretical and empirical scrutiny.

Although there have been many debates about the appropriateness of different perspectives in the context of evaluating health programs, this is not the focus of this paper. The aims of this paper are more modest, and the question we hope to answer is whether methodological guidance with respect to the inclusion of future unrelated medical costs is appropriate within a healthcare perspective. The structure of this paper is as follows. First, we show the role of future unrelated medical costs in a theoretical model describing the healthcare perspective. Then, we will illustrate the practical relevance of future unrelated medical costs using a recently published economic evaluation on transcatheter aortic valve implantation (TAVI) that has been produced using NICE methodological guidance (Watt *et al.*, 2012). In our discussion section, we return to the issue of methodological guidance.

2. THEORETICAL MODEL

Recently, optimal decision rules within a healthcare perspective in which healthcare budgets are exogenously set periodically have been put clearly forward (Claxton *et al.*, 2010, 2011). Table I (adapted from Claxton *et al.*, 2011) displays the incremental costs and health gains of intervention i in two periods expressed in current health. For simplicity, it is assumed that discount rates are equal for costs and quality-adjusted life years (QALYs) throughout this paper (relaxing this assumption does not alter the conclusions in any way).

If the healthcare budget is fixed, all opportunity costs fall within the healthcare sector. In this case, the threshold represents the incremental cost-effectiveness ratio (ICER) of the displaced healthcare programs in that period (i.e., the least cost-effective program that was still funded). As the budget is set periodically, the

Table I. Incremental costs and health gains in two periods for intervention *i* expressed in current health

| | Period | |
|---------------------------------|------------------|-------------------------|
| | 1 | 2 |
| Present value of health gained | Δh_1 | $\Delta h_1/(1+r)$ |
| Present value of health forgone | $\Delta c_1/k_1$ | $\Delta c_2/(k_2(1+r))$ |

Δh_1 and Δh_2 denote incremental quality-adjusted life years in periods 1 and 2; Δc_1 and Δc_2 denote incremental healthcare expenditures in periods 1 and 2; k_1 and k_2 denote the cost-effectiveness threshold in periods 1 and 2; and r denotes the discount rate.

threshold may change over time. According to the decision rules within a healthcare perspective, an intervention should be adopted if the present value of health gained owing to implementing intervention *i* outweighs the present value of health foregone (the health benefits of the displaced activities):

$$\Delta h_1 + \frac{\Delta h_2}{(1+r)} > \frac{\Delta c_1}{k_1} + \frac{\Delta c_2}{k_2(1+r)} \tag{1}$$

Equation (1) can be rearranged to obtain the decision rule that costs per QALY need to be lower than the current threshold in order to be adopted:

$$\frac{\Delta c_1 + \frac{k_1}{k_2} \frac{\Delta c_2}{(1+r)}}{\Delta h_1 + \frac{\Delta h_2}{(1+r)}} < k_1 \tag{2}$$

In case the threshold is equal over time or when the changing value of the threshold is accounted for by differential discounting of costs and effects, the term k_1/k_2 can be dropped from Equation (2). For notational simplicity and without any loss of generality, we will assume in what follows that the threshold is constant over time.

To illustrate the role of future related and unrelated costs within this framework, let us suppose there are two interventions (the already-mentioned intervention *i* and intervention *j*). We want to assess the optimal mix of these interventions over 2 years. Intervention *i* is targeted at disease *x*, and intervention *j* is targeted at disease *y*. One can think of intervention *i* as for instance an intervention targeted at cardiovascular disease and intervention *j* as a treatment for low-back pain. From the perspective of intervention *i*, all spending on intervention *j* can be seen as unrelated and vice versa. Interventions *i* and *j* both increase quality of life and costs in each period conditional on being alive (denoted *Q* and *C*). Next to quality of life and costs, intervention *i* also increases length of life by influencing the survival probability to year 2 denoted by *S*:

$$h_1 = Q_1(i, j); h_2 = S(i)Q_2(i, j); c_1 = C_1(i, j); c_2 = S(i)C_2(i, j) \tag{3}$$

h_1 and h_2 denote total QALYs in years 1 and 2; c_1 and c_2 denote total healthcare expenditures in years 1 and 2. In accordance with previous literature (Garber and Phelps, 1997), we will assume that the cross partial derivatives of the cost functions are zero ($\partial^2 C_1/\partial i \partial j = 0; \partial^2 C_2/\partial i \partial j = 0$) so that, conditional on survival, costs of intervention *i* are not influenced by intervention *j* and vice versa. We will define a Lagrangian function *L* with B_1 and B_2 denoting the total healthcare budget for the two periods to investigate conditions for optimal decision making in this framework (β acts as the time preference discount factor $(1+r)^{-1}$):

$$\text{Max } L = Q_1(i, j) + \beta S(i)Q_2(i, j) + \lambda_1 [B_1 - C_1(i, j)] + \lambda_2 \beta [B_2 - S(i)C_2(i, j)] \tag{4}$$

Equation (4) illustrates that if survival increases because of intervention *i*, less can be spent per person in the second period as the budget in period 2 is set exogenously and thus is not influenced by survival.² Fixity of the

²This differs from previous analyses. For instance, Garber and Phelps (1997) implicitly assumed a per-patient period-specific budget constraint.

healthcare budget is a crucial assumption underlying this operationalization of the healthcare perspective. It implies that the total healthcare budget cannot be influenced by the decision-making body (so that all included interventions must have opportunity costs in terms of displaced activities) nor by medical interventions through their impact on population size. The budget is taken as a given. First-order conditions with respect to interventions i and j are

$$\frac{\partial L}{\partial i} = 0 \rightarrow \frac{\partial Q_1}{\partial i} + \frac{\partial Q_2}{\partial i} \beta S(i) + \frac{\partial S}{\partial i} \beta Q_2(i, j) = \lambda_1 \frac{\partial C_1}{\partial i} + \lambda_2 \beta \left[\frac{\partial C_2}{\partial i} S(i) + \frac{\partial S}{\partial i} C_2(i, j) \right] \quad (5)$$

$$\frac{\partial L}{\partial j} = 0 \rightarrow \frac{\partial Q_1}{\partial j} + \frac{\partial Q_2}{\partial j} \beta S(i) = \lambda_1 \frac{\partial C_1}{\partial j} + \lambda_2 \beta \frac{\partial C_2}{\partial j} S(i) \quad (6)$$

The first-order conditions (5) and (6) clearly illustrate the difference between life-prolonging interventions (intervention i) and interventions that only increase the quality of life (intervention j). Both costs and benefits of unrelated medical care appear in the first-order condition of intervention i , and the weight they have is determined by the incremental effect of intervention i on survival. However, costs of unrelated medical care do not appear in the first-order condition of intervention j . First-order conditions (5) and (6) can be rearranged to obtain ICERs for interventions i and j if we make use of the fact that λ_1 and λ_2 can be interpreted as the number of QALYs that can be obtained by increasing the budget in year 1 or 2.³ These shadow prices equal the inverse of the ICER of the displaced activities within the healthcare budget in the two periods ($k_1 = 1/\lambda_1$ and $k_2 = 1/\lambda_2$). Assuming that the threshold is equal in two periods ($k = 1/\lambda_1 = 1/\lambda_2$), we can reformulate the first-order conditions for optimality as

$$\frac{\frac{\partial C_1}{\partial i} + \frac{\partial C_2}{\partial i} \beta S(i) + \frac{\partial S}{\partial i} \beta C_2(i, j)}{\frac{\partial Q_1}{\partial i} + \frac{\partial Q_2}{\partial i} \beta S(i) + \frac{\partial S}{\partial i} \beta Q_2(i, j)} = k \quad (7)$$

$$\frac{\frac{\partial C_1}{\partial j} + \frac{\partial C_2}{\partial j} \beta S(i)}{\frac{\partial Q_1}{\partial j} + \frac{\partial Q_2}{\partial j} \beta S(i)} = k \quad (8)$$

Equations (7) and (8) state that interventions i and j should be expanded up until the point their marginal cost-effectiveness equals the threshold. Equation (8) demonstrates that costs and benefits of unrelated medical care do not influence the ICER of interventions that do not extend life (note that total health gains and budget impact are influenced by intervention j). Equation (7) demonstrates that both costs and benefits of unrelated medical care (costs and benefits of intervention j in this example) need to be taken into account when calculating the ICER of life-prolonging interventions.

Inspection of Equation (7) reveals some interesting observations. First of all, the ratio of the third terms in both the numerator and denominator ($(\partial S/\partial i C_2(i, j))/(\partial S/\partial i Q_2(i, j))$) can be interpreted as the average cost-effectiveness of all medical care in that period. Furthermore, Equation (7) illustrates that an exogenous decrease in the costs of intervention j in period 2 implies an increase of both interventions i and j . Finally, it is worth noting that period-specific budget constraints do not imply that costs of unrelated medical care can be safely ignored, which has been claimed previously (Lee, 2008). Crucial is the fact that, although survival probabilities may alter medical spending, they do not influence the healthcare budget. The fact that costs for a specific program conditional on survival are independent of other programs ($\partial^2 C_2/\partial i \partial j = 0$) is

³This is a well-known result for optimization problems solved using Lagrange multipliers (e.g., Chiang, 1984). In these optimization problems, the value of the Lagrange multiplier at the solution of the problem equals the rate of change in the maximal value of the objective function as the constraint is relaxed. As the objective is to maximize QALYs, in our model, λ_1 and λ_2 can be interpreted as the amount of QALYs that can be obtained if the budget constraints are relaxed.

irrelevant in this context as total spending on programs is influenced by other programs through its influence on survival ($\partial^2 c_2 / \partial i \partial j = (\partial C_2 / \partial j)(\partial S / \partial i)$). Note that Equations (7) and (8) can also be derived if the threshold varies over time or if one also allows for differential discounting of costs and effects (Claxton *et al.*, 2011).

Now, let us return to the decision rule as proposed in Equation (1). Equation (1) is similar to the first-order condition (5) as $\Delta h_1 \approx \partial Q_1 / \partial i$; $\Delta h_2 \approx (\partial Q_2 / \partial i)S(i) + (\partial S / \partial i)Q_2(i, j)$ and $\Delta c_1 / k_1 \approx \lambda_1(\partial C_1 / \partial i)$; $\Delta c_2 / k_2 \approx \lambda_2[(\partial C_2 / \partial i)S(i) + (\partial S / \partial i)C_2(i, j)]$.

By making a distinction between disease x and disease y , we can decompose total spending $c_t = c_t^x + c_t^y$. As intervention i is targeted at disease x , spending in period 1 on disease y is unaffected by intervention i , so $\Delta c_1 = \Delta c_1^x$. However, because survival is affected by intervention i , there will also be costs for disease y in added life years: $\Delta c_2 = \Delta c_2^x + \Delta c_2^y$. Δc_2^y are the so-called future unrelated medical costs as they are purely the result of living longer. By making a distinction between diseases x and y , Equations (1) and (2) can be rewritten as

$$\Delta h_1 + \frac{\Delta h_2}{(1+r)} > \frac{\Delta c_1}{k} + \frac{\Delta c_2^x}{k(1+r)} + \frac{\Delta c_2^y}{k(1+r)} \tag{9}$$

$$\frac{\Delta c_1 + \Delta c_2^x(1+r)^{-1} + \Delta c_2^y(1+r)^{-1}}{\Delta h_1 + \Delta h_2(1+r)^{-1}} < k \tag{10}$$

From Equations (9) and (10), we can see that only if costs of unrelated medical care are included applying the decision rules leads to an optimal use of the healthcare budget. From Equation (9), we can see that when costs of unrelated medical care are excluded in an economic evaluation of a life-prolonging intervention, this leads to an underestimation of incremental medical expenditures and consequently to an underestimation of the present value of health foregone of $\Delta c_2^y / (k(1+r))$. The higher the Δc_2^y or the lower the k , the bigger the underestimation of the health that would be displaced if the technology is adopted. From Equation (10), we can see that the ICER is underestimated by a factor $\Delta c_2^y(1+r)^{-1} / (\Delta h_1 + \Delta h_2(1+r)^{-1})$ if costs of unrelated medical care are excluded. As only evaluations of interventions that increase length of life are directly affected by this underestimation, this practice results in implicitly favoring life-prolonging treatments over those that improve quality of life. Note also that including the costs of unrelated medical care will have a different effect on the ICER, not just for every therapeutic area but for every pattern of survival payoffs. It is worth noting that these results for medical costs are similar to those found by Meltzer (1997) when analyzing a societal perspective, with the major difference being that future consumption and non-medical expenditures are omitted in a healthcare perspective.

2.1. Should unrelated medical care be evaluated by itself?

One of the informal arguments used to exclude future unrelated medical costs is that the costs of unrelated medical care could be excluded as the costs thereof refer to specific interventions that need to be evaluated themselves (Morris *et al.*, 2007). Here, we will explore this argument and investigate whether this line of reasoning leads to optimal decisions. Following this argument, the benefits of future unrelated medical care obviously should also be excluded. Thus, the health gains of a life-prolonging intervention should be corrected for the effects of unrelated medical care. For the sake of argument, let us assume in our model as described earlier that the health gains in period 2 can be written as the sum of the effects of related and unrelated care, $\Delta h_2 = \Delta h_2^x + \Delta h_2^y$, and that $\Delta h_1 = \Delta h_1^x$.⁴ A way to think about this is that spending on

⁴This assumption will often be violated in practice as there are interactions between treatments in terms of both medical expenditures and health gains.

disease y in period 2 increases quality of life with some amount independent of the spending on disease x ($\partial^2 Q_2 / \partial i \partial j = 0$). Equation (9) then becomes

$$\Delta h_1 + \frac{\Delta h_2^x}{(1+r)} + \frac{\Delta h_2^y}{(1+r)} > \frac{\Delta c_1}{k} + \frac{\Delta c_2^x}{k(1+r)} + \frac{\Delta c_2^y}{k(1+r)} \quad (11)$$

From this equation, we can see that ignoring the costs and benefits of unrelated medical care only results in the same decisions if $\Delta h_2^y = \Delta c_2^y / k$. This would imply that only if the cost-effectiveness of unrelated medical care equals the threshold ($\Delta c_2^y / \Delta h_2^y = k$) can the costs and benefits of unrelated medical care be safely ignored. However, *a priori*, there is no reason to assume that the cost-effectiveness of unrelated medical care in general equals the threshold. In case they are unequal, both costs and benefits of unrelated medical care should be included in the analyses in order to reach identical decisions as prescribed by Equation (1). A necessary requirement for these decisions to be optimal is that unrelated medical care is itself cost-effective: $\Delta c_2^y / \Delta h_2^y \leq k$. Theoretically, this should be the case as, in the optimum, there should be no care that is cost-ineffective. If we look at Equation (7), including the costs and benefits of unrelated medical care implies including a term that is a sort of average cost-effectiveness of total medical expenditures, which should be by definition below the threshold, suggesting that including costs and benefits of unrelated medical care lowers the ICER in general. If $\Delta c_2^y / \Delta h_2^y < k$, excluding the costs and benefits of unrelated medical care can lead to an inappropriate rejection of an otherwise cost-effective technology as including the costs and benefits of unrelated medical care can never push an intervention above the threshold but may cause an intervention's ICER to decrease below the threshold.

2.2. Broadening the decision context

Normally, when carrying out economic evaluations, the decision context is limited to decisions in one disease area. In our model, this means that when we are deciding whether to implement intervention i , we are not deciding whether to implement intervention j . However, in case $\Delta c_2^y / \Delta h_2^y > k$, this implies we are not at the optimum and that unrelated medical care is not cost-effective. This implies that including the costs and benefits of unrelated medical care may cause an ICER to increase above the threshold. It is then critical to think about the commitments we are making when we are evaluating intervention i targeted at disease x for a certain group. If we are also committing to spending on disease y for that group (even if that is cost-ineffective), both costs and benefits of unrelated medical care need to be included as they represent relevant opportunity costs. If spending on disease y is not a necessary consequence of implementing intervention i , the decision context can be broadened. Consequently, one might explore the cost-effectiveness of unrelated medical care by explicitly modeling the provision of unrelated medical care by breaking down the analyses into two different interventions that we are evaluating simultaneously. One intervention relates to the costs and benefits of both related and unrelated care (intervention i), and one relates to the costs and benefits of related care only (let us call this intervention i'). Performing such an analysis would require constructing a counterfactual without the provision of unrelated medical care. If this counterfactual can be constructed, we can apply the basic decision rules of cost-effectiveness analyses to decide whether intervention j (unrelated medical care) should be implemented for this patient group. If $\text{ICER}(i') > \text{ICER}(i)$, the option without unrelated medical care is ruled out by extended dominance as including the costs and benefits of unrelated medical care decreases the ICER. If $\text{ICER}(i') < \text{ICER}(i)$ compared with $i' < k$, including unrelated medical care increases the ICER, but in total, the program is still cost-effective. In this case, costs of unrelated medical care need to be included as they represent opportunity costs that will not be taken into account in any other economic evaluation because without the intervention, the relevant persons would not survive to the future periods. Only if intervention i as compared incrementally with i' shows that unrelated medical care for this particular patient group is cost-ineffective ($\text{ICER}(i$ compared with $i') > k$) is it optimal not to commit

to spending on unrelated medical care for that particular patient group. This then implies that the life-prolonging intervention could be provided to patients, but they would need to be excluded from unrelated future spending. It seems difficult to envisage how this could be acceptably introduced in practice.

3. PRACTICAL RELEVANCE: TRANSCATHETER AORTIC VALVE IMPLANTATION

As a proof of concept of how to estimate future unrelated medical costs and to illustrate the practical relevance of including future unrelated medical costs within a healthcare perspective, we will use the example of TAVI. TAVI, as compared with standard therapy, significantly reduces the rates of death from any cause for patients with severe aortic stenosis and coexisting conditions, who are not candidates for surgical replacement of the aortic valve (Leon *et al.*, 2010). Recently, an economic evaluation of TAVI for the English population was produced using NICE guidance. This resulted in a base case ICER estimate of £16,100 per QALY (Watt *et al.*, 2012). Given that this ICER is just below the threshold commonly used by NICE, TAVI may well be adopted in the NHS. Watt *et al.*, in their evaluation of TAVI, used modeling techniques to extrapolate survival beyond follow-up of the TAVI trial but only included medical costs related to the treatment of heart disease. In line with NICE guidelines, costs of reoperation were included, but costs of other diseases in life years gained, such as dementia, were excluded. In case of TAVI, eligible patients are typically old (>75 years), face a high mortality risk, and generally have multiple comorbidities (Leon *et al.*, 2010). Healthcare consumption for other diseases besides heart disease in additional life years of this group of patients can therefore be expected to be relevant (Yang *et al.*, 2003; de Meijer *et al.*, 2011). Hence, a substantial impact on incremental cost-effectiveness of TAVI is expected when the medical costs thereof are included.

To estimate the influence of unrelated medical costs on the cost-effectiveness of TAVI, we estimated future unrelated medical costs per QALY for the TAVI intervention and added this estimate to the original ICER. The general framework we employed to calculate future unrelated costs for the TAVI intervention in the English context is similar as that employed in a recently developed tool that allows estimation of future unrelated medical costs in a standardized manner for the Netherlands, called PAID 1.0 (van Baal *et al.*, 2011). This means that lifetime healthcare expenditures for unrelated diseases can be written as the sum of per-capita disease-specific expenditures that vary by age and proximity to death:

$$c(n)^y = \sum_a \sum_{i \notin x}^n c_{a,n-a}^i \quad (12)$$

where $c(n)^y$ denote lifetime unrelated medical costs for somebody who dies at age n , a denotes current age ($n - a$ thus indicates proximity to death), and i is an index that runs over all possible diseases except for the set of related diseases denoted x (in the TAVI example, cardiovascular disease). Equation (12) can be thought of as lifetime health expenditures for unrelated diseases if current health expenditure pattern as observed in the general population were to remain constant.⁵ Expectations over n at different ages (proportion alive at a certain age) for the TAVI group as well as the comparator group were taken from the survival curves as used by Watt *et al.* (2012) and allowed us to calculate incremental future unrelated medical costs Δc^y for the TAVI intervention using Equation (12). The rationale behind Equation (12) is that rather than underlying disease status, age and time to death act as proxies for disease status to estimate future unrelated medical costs.

To estimate future unrelated medical costs per capita stratified by age and time to death $\left(\sum_{i \notin x} c_{a,n-a}^i\right)$, we used several data sources and proceeded in two steps. First, we constructed estimates of total medical costs by age and time to death. In the second step, these estimates were corrected for the costs of related diseases

⁵It should be noted that Equation (12) implicitly assumes that costs of diseases, conditional on age and time to death, are independent. More importantly, it also implies that unrelated medical costs are independent of related medical costs. This assumption, of course, may be violated in practice. In the discussion section, we elaborate on this issue.

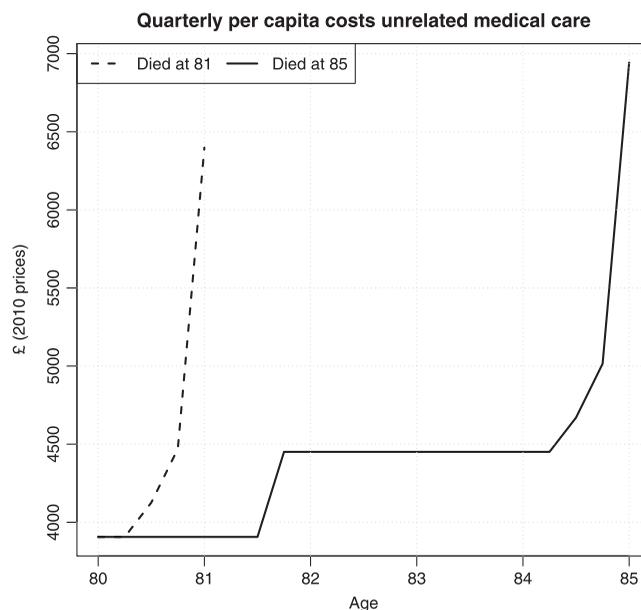


Figure 1. Estimates of per-capita National Health Service and Personal and Social Services costs by age and time to death (2010 price level)

that were already taken into account by Watt *et al.* For total hospital expenditures, we used estimates stratified by proximity to death as published by Seshamani and Gray (2004a). For other NHS expenditures, we used non-hospital per-capita expenditures as estimated by the Office of Health Economics (Hawe *et al.*, 2011). Non-hospital expenditures were not assumed to depend on proximity to death but rather on age only, which seems a reasonable assumption considering that the average age of someone receiving the TAVI intervention is 80 years. As the costs of cardiovascular disease were already included in the study by Watt *et al.*, we multiplied the estimates of total medical costs by age and time to death with 0.9 as the spending on cardiovascular disease amounts to roughly 10% of all NHS expenditures (Hawe *et al.*, 2011). Per-capita PSS costs related to care for the elderly were taken from a recent publication by the Nuffield trust (Wittenberg *et al.*, 2012). As the study by Watt and colleagues did not take into account any PSS costs, we did not need to correct these costs for double counting. All costs were inflated to the year 2010. Figure 1 displays the resulting estimates of unrelated medical costs for various ages at death.

From Figure 1, we can see that per-capita unrelated medical costs increase sharply when approaching death. As non-hospital expenditures increase by age, unrelated medical costs are higher conditional on proximity to death for somebody who dies at an older age. If we combine the estimates presented in Figure 1 with the survival curves, incremental costs of future unrelated medical care (Δc^v) per person receiving the TAVI intervention amount to £11,600. With a threshold of £20,000, ignoring £11,600 per patient implies an underestimation of the present value of health foregone of more than 0.5 QALY per patient that received the TAVI intervention. These incremental costs of future unrelated medical care translate into an increase of the ICER of £7400 per QALY. The ICER increases from £16,100 to £23,500 per QALY gained, thereby changing from an ICER that is below the most recently published threshold to an ICER above this threshold (Claxton *et al.*, 2013).

4. DISCUSSION AND CONCLUSION

Current guidelines for health economic evaluations mostly prescribe that future unrelated medical costs should be excluded from the analysis, also when taking a healthcare perspective. In our theoretical section,

we showed that optimal decisions within a healthcare perspective however require future unrelated medical costs to be included. We hypothesized that current guidelines might be the result of more informal arguments mentioned in health economics textbooks (Drummond *et al.*, 2005; Morris *et al.*, 2007). We demonstrated these informal arguments to be invalid, potentially leading to suboptimal decisions. Using TAVI as a case study, we illustrated the feasibility of estimating future unrelated medical costs for the English context. The TAVI case study illustrated that the impact of including future unrelated medical costs need not be negligible and can clearly matter in practice also when taking a healthcare perspective. In our theoretical section, we paid special attention to the informal argument used to exclude the costs of unrelated medical care, which states that costs of unrelated medical care need not be included as unrelated medical care needs to be evaluated itself. This argument is flawed for several reasons. Excluding the costs of unrelated medical care, while implicitly including the benefits thereof, does not lead to an optimal use of the healthcare budget. Consistency requires that the benefits of unrelated medical care should also be excluded if costs thereof are excluded. However, we showed that theoretically both costs and benefits of unrelated medical care need to be included if unrelated medical care in itself is cost-effective. Excluding the costs and benefits in these circumstances may erroneously result in labeling a cost-effective intervention as cost-ineffective. If unrelated medical care is not considered a necessary consequence of the intervention under evaluation, unrelated medical care can be judged on its own merits by broadening the decision context. Also, in that case, one needs to quantify the costs of unrelated medical care when evaluating related medical care.

4.1. Consequences of current guidelines

As interventions that only increase quality of life are not affected by the inclusion or exclusion of costs of unrelated medical care, ignoring future unrelated medical care favors life-prolonging interventions at the expense of technologies that improve quality of life. More importantly, it misrepresents true opportunity costs within a healthcare budget of decisions to start using specific technologies. Although the numerical impact may differ per intervention, in general, one can state that the impact of unrelated medical care is larger for life-saving interventions targeted at elderly and frail people. Ignoring the costs of unrelated medical care also hampers comparisons between different interventions. Hence, given current guidelines, even economic evaluations following the same methodological guidance may be incomparable. If the ICERs of life-prolonging technologies do not include costs of unrelated medical care, then they can only be compared with other technology assessments that included the same set of cost categories (in case of TAVI technologies, aimed at cardiovascular disease). That in some ways resembles the methodology proposed by the Institut für Qualität, Wirtschaftlichkeit, Gesundheit in Germany, which was criticized for only allowing to optimize disease-specific budgets (Brouwer and Rutten, 2010; Sculpher and Claxton, 2010).

4.2. Unrelated medical care and the threshold

In our theoretical model, we assumed that the value of k is set appropriately. However, deriving a value for k is difficult. Until recently, the threshold used by NICE was influenced by technology appraisals submitted to NICE, excluding future unrelated medical costs (Devlin and Parkin, 2004; Appleby *et al.*, 2007; McCabe *et al.*, 2008). If the threshold is based on an inventory of published ICERs, k is underestimated, which demonstrates another mechanism through which ignoring costs of unrelated medical care distorts making optimal decisions. More recently, there have been attempts to estimate the threshold empirically by exploiting regional variations in NHS spending and mortality (Claxton *et al.*, 2013). However, in these analyses, increases in disease-specific spending instead of total spending were linked to changes in longevity, suggesting that also in the estimates of the threshold, unrelated future medical costs were excluded and thus that the threshold may have been underestimated.

4.3. Is unrelated medical care cost-effective?

Given that it might be difficult to tease out the benefits of unrelated medical care, it is important to know whether we can assume that unrelated medical care in general is cost-effective or not. As unrelated medical care consists of a mix of different interventions, the cost-effectiveness of unrelated medical care in many cases will be some sort of an average return to healthcare expenditures. Empirical research in this area indicates that, on average, investments in health care appear to have been very cost-effective (Cutler *et al.*, 2006; Nixon and Ulmann, 2006; Cutler and McClellan, 2011; Heijink *et al.*, 2012). Including both costs and benefits of unrelated medical care makes intuitive sense. For instance, the benefits of cancer screening not only depend on cancer care in a country but also on the care for cardiovascular disease. However, in some instances, unrelated medical care itself is clearly cost-ineffective if judged against conventional thresholds. An example of the latter would be economic evaluation of statins in patients on dialysis (Grima *et al.*, 2012). If higher medical costs are incurred in certain patient groups, the relevant question simply is whether we are willing to sacrifice more resources to yield health gains in that particular group (van Baal *et al.*, 2013). This relates to efficiency and equity concerns, which need to be addressed explicitly.

4.4. Empirical challenges

In our case study, we used different sources to estimate the unrelated medical costs per capita required for our calculations. Instruments to estimate costs of unrelated medical care in a standardized manner exist for the USA (Meltzer *et al.*, 2000), Canada (Manns *et al.*, 2003), Denmark (Kruse *et al.*, 2012), and the Netherlands (van Baal *et al.*, 2011), but not for England and other jurisdictions. Further improvements in this area are needed and should be possible given the data sources available (Martin *et al.*, 2008; Seshamani and Gray, 2004b). Note that age-specific per-capita costs stratified by time to death as observed in the general population do not always suffice. For frail populations, with high levels of comorbidity, these cost estimates might be inappropriate as they underestimate unrelated medical care in those populations. For the TAVI case, we think this might also be the case. Although not much is known about the healthcare use of patients with aortic stenosis, a recent study (Kasteridis *et al.*, 2014) looked at healthcare expenditures for different patient groups, including patients with heart failure.⁶ They found that the average annual cost of a person with heart failure is more than 30% higher than the average person aged 80 years or above. If we would apply this percentage to our estimate of the costs of unrelated medical care, this would mean that the ICER would increase roughly by £10,000 per QALY rather than £7400 per QALY if future unrelated medical costs are included.

4.5. Healthcare perspective

A final remark is in its place with respect to the healthcare perspective, as advocated by numerous decision-making bodies, including NICE. A critical assumption underlying the healthcare perspective is that the size of the healthcare budget is determined independently of the allocation of the healthcare budget. This implies that productivity gains and/or consumption costs as a result of living longer owing to new medical technology not only cannot influence the choice between different technologies but also cannot influence the size of the healthcare budget. This is difficult to defend from a theoretical perspective. Clearly, it is insufficient to simply use the label of extra-welfarism (e.g., Brouwer *et al.*, 2008) in defense of systematically making incompletely informed, and thus potentially welfare-reducing, decisions. By adhering to a strict healthcare perspective, real opportunity costs and gains outside the healthcare sector, caused by medical interventions, are systematically ignored regardless of their size and nature. These opportunity costs may entail both private consumption and publicly funded consumption. In case of the latter, these opportunity costs hit public budgets as does healthcare spending. For the TAVI intervention, broadening the perspective would primarily mean including non-medical

⁶Aortic valve stenosis often leads to heart failure and is a common cause of heart failure.

survivor consumption, as the target group, owing to their age and poor health, will typically not be productive (both inside and outside the market). Including these survivor costs may even have a stronger impact on the cost-effectiveness than including future unrelated medical costs (Johannesson *et al.*, 1997; Manns *et al.*, 2003).

5. CONCLUSION

We argue that optimal decision making within a healthcare perspective requires inclusion of future unrelated medical costs. Hence, guidelines in jurisdictions advocating a healthcare perspective should recommend inclusion of these costs, although these now often explicitly advocate exclusion. In England and Wales, this can be achieved simply by deleting section 5.5.7 of the NICE (2013) Technology Appraisal Methods Guideline. This leaves less room for researchers to exert discretion as to which costs to include in economic evaluations as the definition of what constitutes 'related' is not clear-cut. More importantly, this leads to more transparent economic evaluations that are more informative for decision makers and to a more efficient use of the healthcare budget.

ACKNOWLEDGEMENTS

The authors would like to thank Karl Claxton and Mark Sculpher for their critical comments on draft versions of this paper and inspiring discussions regarding future costs. The contributions of Pieter van Baal and Werner Brouwer were supported by the Network for Studies on Pensions, Aging, and Retirement as part of the project 'Rising life expectancy: causes and consequences in the Netherlands'. David Meltzer is supported by a Mid-Career Career Development Award from the National Institute of Aging (1 K24 AG031326-01, D. Meltzer). The usual disclaimer applies.

REFERENCES

- Appleby J, Devlin N, Parkin D. 2007. NICE's cost effectiveness threshold. *BMJ* **335**; 358–359.
- Brouwer WBF, Rutten FFH. 2010. The efficiency frontier approach to economic evaluation: will it help German policy making? *Health Economics* **19**; 1128–1131.
- Brouwer WB, Culyer AJ, van Exel N, Rutten FF. 2008. Welfarism vs. extra-welfarism. *Journal of Health Economics* **27**; 325–338.
- Claxton K, Paulden M, Gravelle H, Brouwer W, Culyer AJ. 2011. Discounting and decision making in the economic evaluation of health-care technologies. *Health Economics* **20**; 2–15.
- Chiang AC. 1984. *Fundamental Methods of Mathematical Economics*. McGraw-Hill Publishers: Singapore.
- Claxton K, Walker S, Palmer S, Sculpher M. 2010. *Appropriate Perspectives for Health Care Decisions*. Centre for Health Economics: York.
- Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, Devlin N, Smith PC, Sculpher M. 2013. *Methods for the Estimation of the NICE Cost Effectiveness Threshold*. Centre for Health Economics: York.
- Cutler DM, McClellan M. 2011. Is technological change in medicine worth it? *Health Affairs* **20**; 11–29.
- Cutler DM, Rosen AB, Vijan S. 2006. The value of medical spending in the United States, 1960–2000. *New England Journal of Medicine* **355**; 920–927.
- de Meijer C, Koopmanschap M, Uva TB, van Doorslaer E. 2011. Determinants of long-term care spending: age, time to death or disability? *Journal of Health Economics*.
- Devlin N, Parkin D. 2004. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Economics* **13**; 437–452.
- Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. 2005. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press.
- Feenstra TL, van Baal PH, Gandjour A, Brouwer WB. 2008. Future costs in economic evaluation. A comment on Lee. *Journal of Health Economics* **27**; 1645–9; discussion 1650–1651.
- Garber AM, Phelps CE. 1997. Economic foundations of cost-effectiveness analysis. *Journal of Health Economics* **16**; 1–31.

- Grima DT, Bernard LM, Dunn ES, McFarlane PA, Mendelssohn DC. 2012. Cost-effectiveness analysis of therapies for chronic kidney disease patients on dialysis: a case for excluding dialysis costs. *PharmacoEconomics* **30**; 981–989.
- Hawe E, Yuen P, Baillie L. 2011. OHE Guide to UK Health and Health Care Statistics. Office of Health Economics: London.
- Heijink R, Koolman X, Westert GP. 2012. Spending more money, saving more lives? The relationship between avoidable mortality and healthcare spending in 14 countries. *The European Journal of Health Economics* **14**(3); 527–538.
- Johannesson M, Meltzer D, O’Conor RM. 1997. Incorporating future costs in medical cost-effectiveness analysis: implications for the cost-effectiveness of the treatment of hypertension. *Medical Decision Making: An International Journal of the Society for Medical Decision Making* **17**; 382–389.
- Kasteridis P, Street A, Dolman M et al. 2014. The importance of multimorbidity in explaining utilisation and costs across health and social care settings: evidence from south Somerset’s symphony project. Centre for Health Economics, University of York, CHE Research Paper 96.
- Kruse M, Sørensen J, Gyrd-Hansen D. 2012. Future costs in cost-effectiveness analysis: an empirical assessment. *The European Journal of Health Economics* **13**; 63–70.
- Lee RH. 2008. Future costs in cost effectiveness analysis. *Journal of Health Economics* **27**; 809–818.
- Leon MB, Smith CR, Mack M, Miller DC, Moses JW, Svensson LG, Tuzcu EM, Webb JG, Fontana GP, Makkar RR. 2010. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *New England Journal of Medicine* **363**; 1597–1607.
- Manns B, Meltzer D, Taub K, Donaldson C. 2003. Illustrating the impact of including future costs in economic evaluations: an application to end-stage renal disease care. *Health Economics* **12**; 949–958.
- Martin S, Rice N, Smith PC. 2008. Does health care spending improve health outcomes? Evidence from English programme budgeting data. *Journal of Health Economics* **27**; 826–842.
- McCabe C, Claxton K, Culyer AJ. 2008. The NICE cost-effectiveness threshold: what it is and what that means. *PharmacoEconomics* **26**; 733–744.
- Meltzer D. 2008. Response to ‘Future costs and the future of cost-effectiveness analysis’. *Journal of Health Economics* **27**; 822–825.
- Meltzer D. 1997. Accounting for future costs in medical cost-effectiveness analysis. *Journal of Health Economics* **16**; 33–64.
- Meltzer D, Egleston B, Stoffel D, Dasbach E. 2000. Effect of future costs on cost-effectiveness of medical interventions among young adults: the example of intensive therapy for type 1 diabetes mellitus. *Medical Care* **38**; 679–685.
- Meltzer D. 2006. Chapter 42. Future costs in medical cost-effectiveness analysis. In *The Elgar Companion to Health Economics*, Jones AM (ed). Edward Elgar Publishing Inc.: Cheltenham, UK.
- Morris S, Devlin N, Parkin D. 2007. *Economic Analysis in Health Care*. John Wiley & Sons: Chichester.
- National Institute for Health and Clinical Excellence (NICE). 2013. *Guide to the Methods of Technology Appraisal*. NICE: London.
- Nixon J, Ulmann P. 2006. The relationship between health care expenditure and health outcomes. *The European Journal of Health Economics* **7**; 7–18.
- Rappange DR, van Baal PH, van Exel NJ, Feenstra TL, Rutten FF, Brouwer WB. 2008. Unrelated medical costs in life-years gained: should they be included in economic evaluations of healthcare interventions? *PharmacoEconomics* **26**; 815–830.
- Sculpher M, Claxton K. 2010. Sins of omission and obfuscation: IQWiG’s guidelines on economic evaluation methods. *Health Economics* **19**; 1132–1136.
- Seshamani M, Gray A. 2004a. Ageing and health-care expenditure: the red herring argument revisited. *Health Economics* **13**; 303–314.
- Seshamani M, Gray AM. 2004b. A longitudinal study of the effects of age and time to death on hospital costs. *Journal of Health Economics* **23**; 217–235.
- Tarn T, Smith M. 2004. Pharmacoeconomic guidelines around the world. *ISPOR Connections* **10**(4): 5–12 [on Line].
- van Baal PHM, Wong A, Slobbe LCJ, Polder JJ, Brouwer WBF, de Wit GA. 2011. Standardizing the inclusion of indirect medical costs in economic evaluations. *PharmacoEconomics* **29**; 175–187.
- van Baal P, Meltzer D, Brouwer W. 2013. Pharmacoeconomic guidelines should prescribe inclusion of indirect medical costs! A response to Grima et al. *PharmacoEconomics* **31**; 1–5.
- Watt M, Mealing S, Eaton J, Piazza N, Moat N, Brasseur P, Palmer S, Sculpher M. 2012. Cost-effectiveness of transcatheter aortic valve replacement in patients ineligible for conventional aortic valve replacement. *Heart* **98**; 370–376.
- Wittenberg R, Hu B, Comas-Herrera A, Fernandez JL. 2012. *Care for Older People*. Nuffield Trust: London.
- Yang Z, Norton EC, Stearns SC. 2003. Longevity and health care expenditures: the real reasons older people spend more. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences* **58**; S2–10.