Time and time costs in economic evaluation: taking a societal perspective

Time and time costs in economic evaluation: taking a societal perspective

Tijd en tijdskosten in economische evaluaties vanuit een maatschappelijk perspectief

Proefschrift

Ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
prof. dr. P.W.C. Akkermans M.A. en
volgens besluit van het college voor
promoties.

De openbare verdediging zal plaatsvinden
op vrijdag 3 december 1999 om 13.30 uur

door

Wubbe Berend Frederik Brouwer

geboren te Stadskanaal

Promotiecommissie

Promotor:	*Prof. dr. F.F.H. Rutten*
Overige leden:	*Prof. dr. E.K.A. van Doorslaer*
	Prof. dr. M.F. Drummond
	Prof. dr. L. Koopmans
	Dr. M.A. Koopmanschap (tevens co-promotor)

ISBN: 90-9013268-6

Printed by: Ridderprint, Ridderkerk

In memory of Jaap van Dijk

Publications

The chapters in this thesis are based on the following publications[a]:

Chapter 2: *Brouwer WBF and MA Koopmanschap, 1999, On the economic foundations of CEA. Ladies and gentlemen, take your positions, Journal of Health Economics, forthcoming*

Chapter 3: *Brouwer WBF, MA Koopmanschap and FFH Rutten, 1998a, Patient and informal caregiver time in cost-effectiveness analysis. A response to recommendations of the U.S. Panel, International Journal of Technology Assessment in Health Care 14, 505-513*

Chapter 4: *Brouwer WBF, MA Koopmanschap and FFH Rutten, 1997a, Productivity costs measurement through quality of life? A response to the recommendations of the U.S. Panel, Health Economics 6, 253-259*

Chapter 5: *Brouwer WBF, MA Koopmanschap and FFH Rutten, 1997b, Productivity costs in cost-effectiveness analysis: numerator or denominator: a further discussion, Health Economics 6, 511-514*

and *Brouwer WBF and MA Koopmanschap, 1998, How to calculate indirect costs in economic evaluations, PharmacoEconomics 13, 563-566*

Chapter 6: *Brouwer WBF, MA Koopmanschap and FFH Rutten, 1999, Productivity losses without absence; measurement validation and empirical evidence, Health Policy 48, 13-27*

Chapter 7: *Brouwer WBF, NJA van Exel, MA Koopmanschap and FFH Rutten, 1999, The valuation of informal care in economic appraisal: a consideration of individual choice and societal costs of time. International Journal of Technology Assessment in Health Care 15, 147-160*

Chapter 8: *Brouwer WBF and FT Schut, 1999, Priority care for employees: a blessing in disguise? Health Economics 8, 65-73*

Chapter 9: *Brouwer WBF, BA van Hout and FFH Rutten, 1999, A fair approach to discounting future effects: taking a societal perspective, Journal of Health Services Research and Policy, forthcoming*

---

[a] Reprinted with permission of Cambridge University Press (chapters 3 and 7), John Wiley & Sons, Ltd (chapters 4, 5a and 8), Adis International, Ltd (Chapter 5b) and Elsevier Science Ireland, Ltd (Chapter 6).

Contents

1. Introduction

1.1 Introduction

Sen (1995) writes:

'While Aristotle agreed with Agathon that even God could not change the past, he did think that the future was ours to make - by basing our choices on reasoning.'

Indeed, one may argue that it is this urge for basing choices on reasoning and a clear analytical framework that has caused the increased attention for economic evaluation of health care services over the last decades. The growing medical possibilities for curing diseases and improving quality of life calls for choices in the health care sector. It is financially impossible to perform all possible health care programs with the budgets allocated to health care, even though the latter mostly have increased over the years, in absolute terms and often also as percentage of gross national income.

Economic evaluations of health care have partially met the demand for a comprehensive analytical tool for appraising different health care programs. Different types of economic evaluation prevail, like *cost-minimization analysis* which is aimed at finding that program that is least costly without relating these costs to outcome, *cost-consequence analysis* which makes a very broad comparison between costs and outcomes, as the latter may be measured in many different outcome units within one study leaving the weighting and aggregation to the user of the study (CCOHTA, 1997). *Cost-benefit analysis* aims to balance the costs and the effects of a certain program by expressing costs and all effects in monetary units, thus assigning monetary values to improvements in quality and length of life. Over the last decades, *cost-effectiveness analysis* (CEA) or *cost-utility analysis* (CUA) have gained in popularity. This type of analysis will be discussed in more detail below, as it is this type of analysis for which most of the recommendations made further on in this thesis apply.

CEA provides decision-makers in the health care sector with a rationale for deciding whether or not to introduce or fund a certain health care program and for the critical appraisal of prevailing health programs. This is normally performed by comparing the costs related to a program with the health outcomes it produces. That way, programs that yield much 'health' per unit of cost may be considered to have a higher priority in receiving funding than those that yield less. In the end, this method should provide decision-makers with a tool that may lead to the maximum of health effects from a given budget. Increasingly over the last decades these health effects are captured in terms of quality adjusted life-years (QALYs), which is a common measure of preference for (changes in) health states (e.g. Weinstein and Stason, 1976; Boyle et al., 1983), incorporating both changes in duration of life and changes in the quality of life in a general way. This type of analysis is sometimes called *cost-utility analysis* to distinguish it from analyses using natural units as outcome measure. Here, in principle the term CEA will also be used for analyses using QALYs as outcome measure.

Although the general idea behind cost-effectiveness analysis as outlined above seems rather straight forward, the practical operationalization of CEA has led to numerous debates and problems. Not only the outcome measurement was discussed, e.g. whether to use and how to view and measure QALYs (e.g. Pliskin et al, 1980; Torrance, 1986; Mehrez and Gafni, 1989; EuroQol Group, 1990), how to weigh QALYs for different groups in society (Harris, 1987; Daniels, 1993; Williams, 1988; Lewis and Charney, 1989), whether to use HYEs (e.g. Mehrez and Gafni, 1989; Bleichrodt, 1995) or DALYs (e.g. Murray, 1994), but also the methodology of cost estimations led to considerable debate (e.g. Gold et al., 1996; Drummond et al., 1997).

This thesis focuses mainly on the valuation of changes in time use within cost-effectiveness analysis, taking a societal perspective on CEA. The valuation of changes in productivity and time use in economic evaluations is an important area of debate. Here, there is a focus on how to value changes in different time uses (paid work, unpaid work and leisure) and how *not* to value these changes. Also, the time of informal caregivers is considered. Two additional subjects related to time are discussed in this thesis: priority care for employees and the discounting of health effects in economic evaluations.

1.2 Areas of debate

Societal perspective
Although economic evaluations of health care assess costs and effects related to a specific health care intervention, the costs included in the analysis are not necessarily exclusively health care costs. Similarly, the health effects incorporated in the analysis do not necessarily have to encompass merely the intended effects, but also unwanted side-effects or effects in persons that are not the primary focus of the intervention should be taken into account. This broad view on both costs and health effects is normally denoted as the societal perspective.

'*When a CEA is conducted from the societal perspective, the analyst considers everyone affected by the intervention and counts all significant health outcomes and costs that flow from it, regardless of who experiences the outcomes or costs.*' (Gold et al., 1996, page 6)

This way all relevant information that may be of interest for the decision-making process is incorporated in the analysis. Although Drummond (1993) rightfully questions whether there is a societal decision-maker who indeed adapts to such a broad societal viewpoint rather than a partial interest (e.g. only the health care budget implications), it is considered important not to confine an economic evaluation to a narrow perspective. Consistently following the chosen societal perspective throughout the analysis and in the applied methods is of importance as well.

One implication of taking a societal perspective is that it should ensure that impacts of interventions on all individuals will receive fair weight. The Rawlsian veil of ignorance, or deciding on distributions of goods and services or decision-making processes under the hypothetical assumption that we are deciding on these matters before we are born, or are aware of when or where we will be born, what capacities we will have, what our health status will be et cetera, may be considered relevant in these matters (e.g. Rawls, 1971; Daniels, 1985; Daniels, 1988; Eddy, 1991; Gold et al., 1996). Using this concept of deciding on distributional processes before we actually are born into this world will at least ensure that all relevant changes are considered, since one is not sure whether one is affected by such a change. Also, even though a certain change may not occur within the sector under study (the health care sector), it will be considered as it will affect certain groups in society. Taking such a

broad societal perspective may be considered to have implications for the method used for adapting health effects for the time at which they occur (chapter 9). For costs, it is important to note that taking a societal perspective means that one 'never counts as a gain what is really somebody else's loss' (Gold et al., 1996). This balancing of losses and gains on a societal level is of utmost importance in determining productivity costs for instance.

General view

A second important subject is the way CEA is viewed by the analyst. Two main streams are the welfarist view and the decision-maker's approach or extra-welfarist view. As Hurley (1998) writes:

'The welfare economic framework, which is familiar to most economists, rests squarely on notions of individual utility or preference as the foundation for analysis. For applied analysis, the empirical approach derived from the welfare economic framework is cost-benefit analysis, a method in which the costs and benefits of the alternatives being evaluated are measured in commensurate units, usually money.'

For analysis of health care services, however, even within a welfare economic framework cost-effectiveness analysis to some extent may be seen as an alternative for cost-benefit analysis. This especially holds if QALYs are interpreted as utilities, although one may have some objections against this interpretation (see chapter 2). The 'extra welfarist approach' was first advocated by Sen (e.g. Sen, 1985), who insists that the traditional welfare economic framework limits the possibilities for looking at specific types of outcome, which he calls 'non-utility information', like equity considerations and freedom of action. Culyer further developed this extra-welfarist approach (Culyer, 1989; Culyer, 1990). Culyer suggests that QALY-weights do not necessarily have to stem from individual utility values and that economic evaluation may be seen as maximizing health effects from a given budget. The emphasis on health rather than utility constitutes a divergence from welfare economics.

Here, the viewpoint is advocated that preference based health outcomes are indeed a quite suitable maximand with a health budget, without excluding non-health outcomes from the analysis. Although, in theory one may hold it possible to fully outline societal preferences, including those on equity, and how individual utility contributes to societal welfare, the current operationalizations of the traditional welfare economic framework seem

unable to fully measure and incorporate all relevant information and considerations in the analysis. Primarily focusing on health instead of on pure utility immediately constitutes a divergence from welfare economics. In this thesis cost-effectiveness analysis is considered a pragmatic tool in the process of decision-making, providing the decision-maker with all relevant changes in costs and effects regardless of where (or when) they occur in society, but often not providing an immediate answer to the question whether or not a certain program should be implemented. This remains the task of the decision-maker, weighing all relevant aspects, also those not captured in terms of a C/E-ratio. In agreement with Culyer it is assumed that the main objective of the health care budget is to produce health effects.

Productivity costs and time costs

The costs related to reduced productivity and impaired ability to perform one's normal activities, paid work unpaid work and leisure are often mentioned as a specific cost category within economic evaluations. Especially the costs related to reduced ability to perform paid work are much discussed (e.g. Gold et al., 1996; Drummond et al., 1987; Drummond et al., 1997; Rutten et al., 1993). Productivity costs or indirect non-medical costs have received a reasonable amount of attention in the literature. There has been debate whether these costs should be incorporated in an analysis (e.g. Gerard and Mooney, 1993; Russell, 1986; McGuire et al., 1988; Hartunian et al., 1981) of health care interventions and especially on how (specific parts of) these costs should be incorporated. It has been stressed that valuing production losses is not equal to valuing lives as such (Fitzgerald and Gafni, 1990; Levin and Jonsson, 1992). Mostly, the methodological debate of measuring and valuing productivity costs focused on absence from paid labor. Traditionally, the human capital method was used. For short-term absence, however, one may wonder whether making up for lost work upon return to work, colleagues taking over certain tasks and canceling non-urgent work will not affect the costs of absence to some degree (e.g. Commonwealth of Australia, 1990; Drummond, 1991; Koopmanschap and Rutten, 1993). In addition, the relationship between changes in production and changes in labor time is less than proportional (De Koning and Tuyl, 1984). For long term absence, the possibility of replacing ill, disabled or prematurely diseased workers may induce a substantial reduction in production losses on a societal level (e.g. Drummond et al., 1987; Williams, 1985; Lindgren, 1981; Levin and Jonsson, 1992; Olsson et al., 1987; Koopmanschap and Ineveld, 1992; Koopmanschap et al., 1995; Koopmanschap and Rutten, 1996b). These concerns led to the

development of the friction cost method (Koopmanschap and Rutten, 1993; Koopmanschap et al., 1995). In this thesis the friction cost method is further advocated as being the most appropriate method from a societal perspective. The use of the friction cost method is embedded in more broad recommendations for valuing patient time (chapter 3). There is also explicit attention for the recent recommendation of the US Panel on valuing productivity costs in terms of quality of life. This recommendation presents a new third way of valuing productivity costs, which will be discussed in chapters 4 and 5.

Productivity costs without absence may also be of importance, especially for certain types of diseases like migraine and dyspepsia (Osterhaus et al., 1992; van Roijen et al., 1995a; Severens et al., 1998). Although from the scarcely available literature on this subject it is clear that this type of productivity loss may constitute a considerable cost in some instances, the exact measurement of the amount of production lost, the degree of compensating mechanisms like colleagues taking over or making up for lost work later and the valuation of the lost production or increased production costs need further attention. Chapter 6 further discusses this subject. The valuation of production losses related to unpaid labor is not without problems either. Although the number of hours of reduced unpaid productivity may be derived through interviews or questionnaires (e.g. Van Roijen et al, 1996), divergent recommendations have been put forward on how to value the reduction in output and possibility to perform unpaid labor (e.g. Drummond et al., 1997; Gold et al., 1996). In chapter 3 some recommendations are made for valuing lost unpaid labor following the opportunity cost principle.

A final matter of debate related to time costs is the impact of health changes on leisure time and role functioning. The fact that people are able to enjoy leisure time in the way that they prefer is a valuable aspect of life and should not be neglected in economic evaluations. However, there is no clear consensus on how to value leisure time in economic evaluations. The same holds for role functioning in paid or unpaid work. In chapter 3 these aspects are discussed and the relationship between productivity costs and quality of life aspects of changed time use is highlighted.

Informal caregivers

Time invested by family and friends in (certain necessary elements of) the treatment of a patient may be an important part of total care in some cases.

Besides costs of giving up on normal time uses, there may also be effects on the wellbeing of the informal caregiver because of fatigue, being restricted in one's social life and so on. In chapter 7 an alternative valuation method of informal care, incorporating the full costs and effects on quality of life of informal care, is presented.

Value of time and priority care

The relationship between direct health care costs and productivity costs became apparent in the Netherlands after the introduction of laws that made firms individually responsible for the continuation of sick pay. A clear tension has arisen between the private responsibility of firms for sick pay and the public responsibility for providing adequate and *timely* care. However, waiting lists are substantial in The Netherlands (Swinkels, 1996). Employers, in collaboration with insurers and hospitals came up with plans for employee-clinics that should provide priority care to employees. These clinics should ensure a swift return to work for employees leading to reductions in production losses. However critics have claimed that non-employees would be disadvantaged by these clinics and that they should therefore be prohibited. In chapter 8 it is shown that this specific intervention can be considered efficient and equitable, demonstrating how investments in health care may lead to gains in other sectors of society. In addition, this relationship between indirect and direct costs *and* between the health care sector and other sectors underlines the need for taking a societal perspective in assessing health care.

Timing of health effects

The discounting of health effects in economic evaluation is a common procedure. Health effects are weighted for the point in time at which they occur. Usually, effects are discounted with a 5% discount rate. The reason for discounting effects is not fully clear and it may be considered quite unfair for future generations or people that will receive health effects from current treatment later on in life, that such a procedure is performed. An alternative view on discounting effects is presented, which may be argued to be more fair. Here, another great philosopher can be quoted. Friedrich Nietzsche (1885) writes that every philosophy is born out of the moral intentions of the writer. Indeed, this may be held true for ethical views on how to discount health effects as well. In chapter 9 discounting health effects from a societal perspective is discussed in depth.

1.3 Questions

From the above discussion numerous questions may be formulated. Some of these questions to be addressed in this thesis are:

- How should the time of patients be valued in economic evaluations of health care?
- What aspects of patient time are to be measured in terms of quality of life and what in terms of costs?
- Can productivity costs be measured reliably in terms of quality of life?
- Are productivity losses at work relevant for economic evaluations and if so, how should they be measured and valued?
- How might one incorporate the impact of the provision of informal care in an economic evaluation of health care?
- Should health effects be weighted for the point in time at which they occur by discounting them and if so, on what basis?
- Can productivity costs related to paid labor provide a rationale for priority treatment of employees?
- Is or should economic evaluation be a strict application of welfare economics? If not, what alternative is there?

These questions are all related to the central theme of this thesis:

How does taking a societal perspective in economic evaluations of health care, following a decision-maker's approach, change methodology, especially those methods concerning time and time costs.

Hopefully, the answers provided in this thesis to the above question will contribute to the ongoing quest for standardization of methodology of economic evaluations and stimulate the discussion on various subjects.

1.4 Outline of this thesis

A first important issue is that one may encounter different views on economic evaluations in general. Where some economists try to embed economic evaluation in welfare economics, others take a perhaps more broad

perspective. These latter economists view economic evaluations as an imperfect tool aiding decision-makers in prioritizing health care expenditures, in which not all standard economic welfare principles may be directly applicable or desirable. As will be further outlined in chapter 2, in this thesis this latter perspective on economic evaluation and especially cost-effectiveness analysis and cost-utility analysis is taken. Aristotle considered economics to be just one of the sciences that could be used in the 'overall-science' of politics. Indeed, the decision-maker's approach may be seen to stand in this tradition, by providing policymaker's with all relevant information, but leaving the final weighting of positive and negative aspects to the policymaker.

Chapter 3 provides an alternative way of valuing patient time in cost-effectiveness analysis, by separating time uses into three categories: paid work, unpaid work and leisure. It is argued that the valuation of time should not be entirely in terms of money or costs, but that there are quality of life aspects to valuing time as well. Mostly, the debate on changes in time use concentrates on a particular part of time costs, the so-called productivity costs.

In chapter 4 the question whether or not productivity costs may be estimated in terms of quality of life as suggested by the US Panel on Cost-Effectiveness in Health and Medicine (Gold et al., 1996) is discussed. If respondents in the valuation of health states are not explicitly instructed *not* to consider any income change resulting from a change in health, these respondents may be assumed to fully take these income changes into account. By further assuming income of individuals a good proxy of the value of their production, also from a societal perspective, one may claim that production losses need not be incorporated in terms of costs in the numerator of the cost-effectiveness ratio (C/E-ratio), but are fully incorporated in terms of health effects in the denominator of the C/E-ratio. Chapter 4 considers whether such an approach may be considered appropriate.

Chapter 5 presents two recent debates in the literature on the measurement of productivity costs. The first discussion was between the 'Erasmus Group' and the US Panel on Cost-Effectiveness in Health and Medicine. After publication of chapter 4, some members of the US Panel responded, mainly criticizing the alternative method for valuing production losses, the friction cost method. Another critique on the friction cost method was provided by Liljas (1998). His main arguments and the responses to them are discussed.

Chapter 6 deals with production losses without absence from paid work, i.e. because of illness at work. This much neglected subject is examined tentatively by comparing three different measurement methods of production losses at work. After a theoretical examination of the different measurement methods, some empirical results are presented from a study of production losses without absence in a sample of employees of a Dutch trade-firm.

In addition to looking at changes in the productivity and time use of patients, there is also attention for the position of informal caregivers, who sacrifice normal time use in order to provide care to a partner, family member or friend. This time represents a scarce resource, and as such should in principle be valued in an economic evaluation. However, as is further explained in chapter 7, the valuation of time of informal caregivers is not without problems. Mostly, one standard valuation is attached to every hour of informal care provided, but one may question the appropriateness of this method.

In chapter 8 priority care for employees as a solution for waiting times is discussed. Employee clinics providing priority care for employees may ensure a substantial reduction in waiting times for employees reducing production losses, but it has to be considered to what extent a distinction in health care provision on non-medical criteria is acceptable. By organizing employee-clinics in such a way that both employees and nonemployees benefit, one may argue that employee clinics are both economically and ethically defendable.

An opinion on how to operationalize a fair representation of future health effects in economic evaluations from a societal perspective is presented in chapter 9. Using the Rawlsian principles of justice and a strong intergenerational viewpoint, one may put forward other recommendations on the discounting of health effects than usually done in the literature.

Chapter 10 concludes this thesis and gives recommendations for future research.

It should be noted that the chapters in this thesis are all based on separate essays for journals. Therefore, all chapters may be read independently and some overlap between various chapters may exist.

2. On the Economic Foundations of Economic Evaluation

Summary

There are still many ongoing debates about several aspects of the methodology of economic evaluations of health care interventions. Some of the disparities in recommendations on methodological issues may be traced back to different viewpoints on CEA in general. Two important views are the welfarist approach, which aims at embedding CEA into traditional welfare economics, and the decision-maker's approach, which takes a broader and more pragmatic view on CEA. The focus in welfarism may be on utility while that of the decision-maker's approach may be considered to be on health. In this chapter it is examined how these two views differ and how these differences may subsequently lead to debates in methodological areas. It is indicated that embedding the practical operationalisation of CEA in welfare economics seems impossible. In a strict welfarist approach it is necessary to view QALYs as being utilities, although one may question whether such an approach to QALYs is appropriate. Also, equity considerations may play an important role in cost-effectiveness analysis and these should preferably be taken into account in a way that reflects societal attitudes towards an equitable distribution of health care. These equity considerations may not always be directly related to utility or efficiency. Furthermore, both camps may prefer different methods for cost measurement in areas such as productivity costs and informal care. A better recognition of the contents and origins of controversies and disputes may enhance the clarity of discussions.

2.1 Introduction

The general methodology of economic evaluation in health care (cost-benefit analysis, but especially cost-effectiveness analysis and cost-utility analysis) and its economic foundations are increasingly becoming a matter of discussion. Recently, in the *Journal of Health Economics* for instance, there was a discussion on economic foundations of cost-effectiveness analysis (CEA) in relation to

the inclusion of unrelated future health care and non-health care costs (Garber and Phelps, 1997; Meltzer, 1997; Weinstein and Manning, 1997). In addition, there is an ongoing debate on several other methodological issues, like whether and how to incorporate productivity costs (e.g., Koopmanschap et al., 1995; Johannesson and Karlsson, 1997, Koopmanschap et al., 1997; Brouwer et al., 1997a; Weinstein et al., 1997; Brouwer et al., 1997b; Brouwer et al., 1998a), how to incorporate equity principles in CEA (e.g., Wagstaff, 1991; Bleichrodt, 1997; Williams, 1997), how to incorporate and value informal care (e.g., Gold et al., 1996; Drummond et al., 1997; Brouwer et al., 1999) and how to discount costs and especially health effects in CEA (e.g., Gold et al., 1996; Gyrd-Hansen and Søgaard, 1998; Van Hout, 1998).

Some of these debates or the lack of consensus on specific subjects may originate from the fact that divergent views on economic evaluation prevail among analysts. Weinstein et al. (1996) write that some of the disparities in the methods that investigators employ '...reflect divergent views on key methodological choices'. Whereas some authors prefer to see economic evaluation as founded in traditional welfare economics, others follow a more pragmatic 'decision-maker's approach' (DMA). Although one might claim that both approaches aim at aiding policymakers in decision-making on health care services, welfarists may try to base methodological recommendations on often individualistic welfare economic models, whereas adherents of the decision-maker's approach base their recommendations on societal values and pragmatic assumptions. Hurley (1998) insists that both lines of thought share shortcomings as well.

Welfarists may criticize the DMA, claiming it has no theoretical framework, as it is not embedded in standard welfare economic theory.[1] The foundation for the methodological choices in the DMA can be viewed as 'maximizing health effects from a given budget'. In addition, the economic foundations of the welfarist approach to economic evaluation are not indisputable either. It may be questioned whether many of the often individualistic, microeconomic theoretical models for determining optimal solutions and recommendations, with their underlying assumptions and specific viewpoints, are fully valid for and instrumental to societal decision-making in the health care sector.

[1] See for instance Johannesson and Karlsson (1997) in their reply to the friction cost method.

In this chapter, we will discuss the decision-maker's approach and the welfarist approach in societal economic evaluations of health care. While the main focus will be on cost-effectiveness analysis, which here includes so-called cost-utility analysis, economic evaluation will be discussed more broadly. It is indicated how these different views may lead to disparities in methodological recommendations. Some practical and theoretical viewpoints will be appraised on their usefulness and consistency. Section 2.2 points out the main differences between the decision-maker's and the welfarist approach. In section 2.3 the divergent views these two approaches have on several methodological issues in economic evaluation are discussed. Finally, section 2.4 gives a conclusion and discussion.

2.2 Competing views on economic evaluation

The general aim and rationale of CEA is to aid decision-making in health care, with the goal of maximizing health benefits from a given budget, taking a societal perspective (Gold et al., 1996). Economic evaluation has come into fashion because it can provide a rationale for choosing certain programs over others, which is convenient when budget constraints do not enable policymakers to implement all possible health care interventions and programs. Taking a societal perspective means that all costs and health effects should be incorporated, regardless of who bears those costs or experiences the health effects. Although there may be a broad consensus on these general outlines of CEA in particular and economic evaluation in general, the subsequent operationalisation of methods shows that different views on economic evaluations may still be encountered. This has been recognized in the literature. For example Drummond et al. (1997) consider three hypothetical analysts with different views on economic evaluation. One wishes to assess in an economic evaluation an *individual's willingness to pay* (WTP) for certain programs, in order to compare these with the costs of a program, performing a cost-benefit analysis[2]. Ultimately, this approach should be consistent with standard welfare economics. The other analysts think of economic evaluations as an *(imperfect) aid to decision-making* regarding the allocation of health care resources. For them willingness to pay is not considered very helpful as it may reflect ability to pay. These latter analysts, following a DMA, are more likely to take a societal perspective which leads them to measure and value 'a wide range of costs and consequences and

[2] See also Hurley, 1998

present them in a way that helps health care decision-makers form a better judgement' (Drummond et al., page 18).

Although the above characterization of welfarism and DMA may be somewhat exaggerated, the general differences outlined nevertheless seem to prevail in many of the ongoing discussions on the methodology of CEA. Some may claim that welfarism in contrast to the DMA can provide guidance in methodological choices. For instance, Weinstein and Manning (1997) state that the DMA or extra-welfarism 'provides little guidance in controversies about what should be counted in the C/E-ratio and how'. However, this view may be questioned. Taking the aim of economic evaluations to be maximizing the number of health effects with a given health care budget while considering other resource changes in society, e.g. through indirect costs, is in itself a frame for analysis and this pragmatic frame may also just as well provide guidance in methodological choices. In addition, Weinstein and Manning (1997) acknowledge that conclusions derived from certain welfare economic models 'may be unsettling to many practitioners of CEA', for instance because ethical considerations overrule basic assumptions of welfare economics.

Welfarism

Let us first consider both lines of thought more carefully. Welfarists have tried to embed economic evaluations into welfare economics. This should ensure that from a welfare economics perspective optimal choices are made in the allocation of health care funds. Also, investments in health care are considered in a framework similar to that for other allocations of scarce resources. Culyer (1991) describes this viewpoint as:

> *'One approach in health economics, which has become the traditional one in economics as a whole, is what Sen (1977a) calls 'welfarist'. This is very much in accord with liberal political opinion and asserts that social welfare ... is a function of only individual welfare (or utility) and judgements about the superiority of one state of the world ... over another are made irrespective of the non-utility aspects of each state. Moreover, the individual welfares (or utilities) are a function only of goods and services consumed'* (page 67).[3]

The economic models often used for embedding economic evaluations theoretically in economic theory are often individualistic micro-economic models. In drawing conclusions from these models for society as a whole, one

[3] Hurley (1998) writes: 'The welfare economic framework ... rests squarely on notions of individual utility or preference as the foundation for analysis'.

of the key assumptions underlying the aggregation of individual results is that society's preference or utility is a perhaps weighted sum of individual's preferences or utility. For instance Loomes and McKenzie (1989) write, in response to Weinstein (1986):

> *'Implicit in Weinstein's approach ... seems to be the view that the values to be used in social decision-making should be some aggregate of individual's values. But this is not the only possible view: for example, the socio-political system may designate certain people to make decisions on society's behalf, and the values these decision-makers use may not be some aggregate of individual's values, but may reflect some other notion of what represents society's overall best interests.* ' (page 94)

Also, strict application of welfarism makes it necessary to view QALYs as utility measures. However, the practice of economic evaluation involves interpersonal comparison and aggregation of these utilities in terms of QALYs, to be able to conclude on a societal level whether or not a program is worthwhile. The question of whether such use of QALYs is fully in line with welfare economics is not easily answered (see section 2.3). A possible solution that does not require the interpretation of QALYs as utilities to be used for aggregation and comparison is the use of contingent valuation, like WTP. This implies performing cost-benefit analysis (CBA) in which health effects are expressed in monetary terms. Pauly (1995) even argues that in general CBA is more appropriate than CEA in the assessment of health care technologies. While within CBA subsequent aggregation and comparison of results is perhaps without problems, the fact that the results reflect ability to pay causes many to be suspicious of the consequent allocation of health care resources, especially since programs may yield health effects in specific socio-economic groups in society. Also, there are serious concerns about the reliability and consistency of WTP-answers (e.g., Pauly, 1995; Johannesson et al., 1996; Krabbe et al., 1997). In addition, there is evidence that the monetary willingness to pay per QALY gained can vary depending on the distribution of wealth and the size of risk-reduction (e.g. Jones-Lee, 1976; Viscusi, 1979; Pratt and Zeckhauser, 1996). A more pragmatic point perhaps in favour of non-monetary outcomes is that policymakers and clinicians may relate less to them than to health outcome measures such as QALYs.

A recent example of an attempt to embed cost-effectiveness analysis in welfare economic theory using a simple individualistic micro-economic utility framework is given by Garber and Phelps (1997). One implication of the

approach of Garber and Phelps is that individual optimality holds when 'at the margin the wage rate is equal to the willingness to pay for an additional unit of time...'. This result implies a strict relationship between productive possibilities and acquiring additional life-years. No explicit societal utility is derived from the mere fact of being alive in the model, or the 'simple' things that make life valuable, like social contact. This type of 'optimality' may be unsettling, both for DMA-adherents and probably most welfarists, since the WTP for less productive time, e.g. of severely handicapped, is consequently low. This demonstrates a gap between these models and real-life valuation of effects in societal decision-making. From a societal perspective, ethical and perhaps also efficiency considerations may lead to opposite conclusions, favoring the health promotion of less fortunate and less productive persons in society. Basing methodological conclusions on models like these should also be performed with caution. It seems important to assess assumptions used in welfare economic models in terms of their ability to reflect the values and judgements endorsed by society or policymakers. In that respect it may be more useful to use social welfare functions rather than individualistic models, incorporating distributional considerations of wealth and health or utility (e.g. Bleichrodt, 1997; Dolan, 1998)[4].

Extra-welfarism

The alternative view on CEA has been put forward by Culyer (1991) and Williams (1993). Culyer calls the alternative view 'extra-welfarism' and indicates that this view includes (or is able to include):

> *'an important class of 'extra' welfare sources ... the non-goods characteristics of individuals (like whether they are happy, out of pain, free to choose, physically mobile, honest). Extra-welfarism thus transcends traditional welfare: it does not exclude individual welfares from the judgements about the social state, but it does supplement*

[4] Note that such models do not fully fall within Culyer's definition of welfarism, as indicated above. Equity weights may imply that individual utility may be 'corrected' or weighted from a societal level. Classical utilitarianism which sets social welfare equal to the sum of individual utilities is abandoned therefore. Also, one could also examine whether other traditional welfare theoretic elements, such as the Pareto-criterion, or even the 'potential Pareto-criterion', may be violated in such models. Moreover, as pointed out by Bleichrodt (1997) one needs an interpersonally comparable utility measure to determine which groups or persons to give priority in allocation of health care resources. As discussed in section 2.3 it is questionable whether QALYs may be interpreted as being such a measure.

> *them with other aspects of individuals (including even the quality of the relationships between individuals, groups and social classes).*

Thus extra-welfarism 'transcends traditional welfare', and does not merely focus on individual utilities. More precisely, within extra-welfarism one does not necessarily have to translate all these aspects into utility-weights in order to take them into consideration in the analysis. As Hurley (1998) notes, extra-welfarism rejects 'the conceptual foundations of the neoclassical framework, particularly the exclusive focus on utility-based notions of welfare.' Referring to Culyer he continues by stating that this framework 'replaces utility with health as the primary outcome of interest for evaluation'.

Indeed, the extra-welfarist view is often interpreted as merely aiming to 'maximize health from a given budget'. Culyer (1989) provides arguments for seeing health outcomes or QALYs less as utilities and more as 'capabilities' brought on by good health. When considering QALYs as representing health capabilities rather than utilities and assigning an equal value to these *capabilities* for all, one could conclude that maximization of these health outcomes or 'health' in general is an appropriate goal in the health care sector. Indeed, the health care budget seems to be particularly aimed at providing individuals a return to good health or to reduce suffering from bad health (while also considering relevant resource changes in other parts of society). The emphasis on health rather than utility may be justified by considering the utility of those suffering from disabilities and diseases. These individuals may perhaps live a life of fulfillment involving a level of utility higher than that of some healthy persons, but still they are entitled to health care in order to minimize their health problems. This demonstrates that the *aim* of providing health care may at best indirectly be utility maximization.

However, extra-welfarism should not be defined as focusing exclusively on health. Culyer (1991) states that the emphasis on health 'is not in principle exclusive, and it seems unlikely that any extra-welfarist would assign zero weights to ... factors only remotely causally linked to health.' (page 96). Therefore, non-health implications may also be incorporated in the analysis and the weights attached to the capabilities of different persons or groups in society may be differentiated for reasons of equity or efficiency. For instance, health effects in elderly may be valued lower than those gained in children (Williams, 1997). By using a broad notion of QALYs as a preference-based health outcome measure that may be corrected for equity and efficiency

considerations (see section 2.3), one might claim to try to *optimize* the production of health benefits from a given health care budget.

Besides more conceptual work like that of Culyer, the development of practical methodological guidelines and instruments for CEA has probably influenced our thinking about CEA as well. The way the field has developed may to some extent be a result of attitudes of for instance policymakers and researchers from different fields towards certain methodological choices. For instance, the way in which quality of life instruments have been constructed, focusing on certain aspects of human life and explicitly or implicitly excluding others, e.g., income changes, probably reflects an existing opinion on what should be measured in terms of quality of life and what not. Moreover, a preference for CEA and non-monetary measures of health effects may be seen as reflecting an implicit rejection of monetarizing health effects by policymakers and analysts.

What we shall call DMA is therefore a pragmatic common sense approach taking as its theoretical framework the aim of optimizing (weighted) health benefits from a given budget. Also, in this approach the evaluation is aimed at *informing* decision-makers rather than *prescribing* what decision should be made or providing a strict ranking of all possible health care interventions. Decision-makers are provided with all relevant information including *non-health* information and then have to make a decision by considering all different aspects. Implicit in the DMA is the general notion that society wishes to maximize a social welfare function. In this respect it is not different from traditional welfare economics.[5] In theory, welfarism may be able to incorporate all relevant information by assigning utility impacts to all relevant changes related to a health care program and specifying a complete and acceptable social welfare function.[6] However, adherents to the DMA may have reservations concerning the possibility to incorporate all relevant information, ethical values and interactions present in society validly in a formal societal welfare function. Also, DMA-adherents may feel *health*, for instance captured in terms of QALYs, to be a better maximand than *utility* with a given health care budget and considering other resource changes and

[5] There may, however, be differences in how this social welfare function is defined and which elements should be included in it.

[6] Again note that this implies an important broadening of the concept of welfarism compared to the definition of Culyer. For instance, it seems necessary to have interpersonally comparable utilities, a complete set of societal valuations of individual utilities, allowing possible violations of the (potential) Pareto-criterion et cetera.

'non-utility' equity considerations, such as considerations of equal access. The manner by which all aspects are weighted and by which a decision is reached is considered to be the task of a policymaker, and weighting by analysts should be aimed at representing societal values[7] and should be made explicit in any economic evaluation. Society may have 'higher goals' than purely following the individual preferences, for example by promoting health care for mentally handicapped or attaching more weight to future effects than individuals do. The inability to fully capture and weight all of these different aspects, their real utility impact and their interactions in fully outlined and defined theoretical terms and decision rules, does not lead to an inability to make decisions. Rather, the policymaker will have to weight different aspects of a program, aided by analysts.

Does it matter which side one takes?

Of course, given the differences in both approaches, one has to ask oneself whether it makes much difference which view on CEA one takes. We feel that different starting points may indeed lead to different methods of analysis and to different conclusions. In the next section, some of these differences and controversies are further discussed. Table 2.1 provides the reader with an overview of some of the areas in which we feel that the difference in viewpoint may lead or has led to differences in methodological choices.

[7] Hurley (1998) insists that '...as a society we have a responsibility to assess systematically the effects of policies in an attempt to ensure that they contribute towards social objectives in ways that reflect social values.'

Table 2.1: Methodological issues in economic evaluation using the welfarist or the decision-maker's approach

Issue	Welfarist approach	Decision-maker's approach
Operationalisation of societal perspective	(Weighted) aggregation of individual preferences in social welfare function	Societal preferences based on an implicit social welfare function
Valuation of health effects	Monetarisation (WTP or a statistical value of a life) or QALYs as utilities	In principle natural units or QALY's (less specific interpretation)
Valuation of time/productivity costs	WTP for time gains/ Human capital method	Human capital method/ Friction costs method
Valuation of informal care	WTP or wage rate (shadow price)	Shadow price or QALY?
Discounting costs	Long term interest rate	Long term interest rate
Discounting health effects	Average of individual time preferences for health	Societal/governmental time preference for health
Equity considerations	Equity-efficiency trade-off is possible, but is a divergence from unweighted individual utility as maximand	May be explicated and weighted separately
Weighting other consequences	Possible, but difficult to translate into utility or 'overruling' individual utility	May be explicated and weighted separately

2.3 Differences in methodology and viewpoint: some examples

QALYs

A first difference between welfarism and the DMA is their view on QALYs. Within the DMA QALYs may be interpreted as an imperfect, yet quite appropriate tool to operationalize health effects. These health benefits are what ultimately should be maximized from a societal perspective using the health care budget, while minding positive and adverse effects on other sectors in society and equity considerations. Culyer (1989) suggests an alternative way to interpret QALYs instead of as utilities. He feels that when using QALYs we should focus more on capabilities that good health brings us rather than on utility, a move away from strict welfarism. Drawing from Sen (1980) Culyer states that

> *'the idea of utility focuses too much on mental and emotional responses to commodities and characteristics of commodities and not enough on what they enable you to do'.* [8]

QALYs as a measure of capabilities brought on by good health then in principle are a good thing to maximize with a health budget. As Broome (1993) states:

> *'I am impressed by a point made by Culyer that using QALYs does not commit one to a narrow - he calls it 'welfarist' - conception of good. QALY analysis assigns values to states of health and leaves it open whether these values are determined by how people feel when in these states, by their preferences about them, or perhaps by some objective principles.'*

However, some preference (or utility) indication or valuation is still needed to see which health (or capability) gain is greater than another. How can we otherwise choose between a program curing one person of hearing difficulties or one person of blindness? Only with some kind of cardinal ranking order is it possible to come to 'the maximand which is a cardinal "utility" index of health' (Culyer, 1991). The fact that we see one health gain as greater than the other shows some societal or other preference or a normative judgement,

[8] One of the problems of focusing on these emotional responses is for instance adaptation (e.g. Hurley, 1998).

rather than anything else. The source of this preference does, however, not necessarily have to be strict individual utility. In terms of Culyer: QALYs may indicate the extent of people's deprivation and health care may be used to remove this.

If we look at the use and interpretation of QALYs within CEA, the attempt to fully embed CEA in welfare economics seems heroic and at this point far from accomplished. The interpretation of QALYs as utilities, needed to firmly embed CEA in welfare economics, is not indisputable if we consider how QALYs are used in CEA, i.e. not only comparing the valuation of health states for one person, but comparing interpersonally and aggregating QALYs. The comparability of utility between persons or groups of persons is problematic, yet for societal decision-making unavoidable. One might differentiate between different types of utility comparison (Sen, 1977b). As indicated by Bleichrodt (1997), when comparing and aggregating QALYs and weighting QALYs/utilities for considerations of equity, one needs to assume cardinal full comparability. Many economists, however, may consider full interpersonal utility comparison off limits. People may rank their own preferences, but since the underlying amount of absolute utility is unavailable, comparing these rankings cardinally between persons seems inappropriate. Although rescaling preferences between 0 and 1 will lead to comparable rankings, obviously this does not necessarily relate to absolute amounts of utility.

Utility measurement for individuals is often based on the theory of expected utility by Von Neumann and Morgenstern (1944). Von Neumann and Morgenstern themselves use utilities in their book as means for exchange and comparison, which might tempt one to extrapolate this possibility of interpersonal comparison to QALYs. However, Von Neumann and Morgenstern in several places in their book assure that they assume a monetary notion of utility ('a quantitative and even monetary notion of utility'), facilitating comparisons and exchange ('substitutable and unrestrictedly transferable between the various players'). Such restrictive assumptions do not apply for QALYs, since these are not interpersonally comparable units in themselves, as the underlying absolute level of utility is unknown. Indeed, Culyer (1989) is correct to assert that an optimistic arthritic may obtain a higher utility level than a healthy Calvinist convinced he is not among the chosen. Rescaling utility into numbers between zero and one hides the absolute levels. Interpersonal comparison and aggregation of QALYs can

therefore not be based solely on the work of Von Neumann and Morgenstern. Interestingly, the monetary notion of utility used by Von Neumann and Morgenstern seems more related to contingent valuation such as willingness to pay and the use of CBA rather than to QALYs and CEA.

The use of two fixed endpoints (death is 0, perfect health is 1) is sometimes indicated as being a rationale for comparing between persons. If one *assumes* that utility is equal for all persons then one may compare the values given by different persons between these two extremes (disregarding states worse than death). One may also interpret this as meaning that the value of health is equal for all in a DMA-like approach. Torrance (1986) for instance indicates that:

> *'The central basis for this method is that the difference in utility between being dead and being healthy is set equal across people. In this way, the method is egalitarian within the health domain; that is, each individual's health is counted equally'.*

By *in principle* attaching equal values to similar health changes across persons or groups, one may add, compare and average scores.[9] Although this procedure of assigning an equal value (in terms of utility) to every person may be considered an ethical way of dealing with health, it effectively constitutes a divergence from welfare economics. Interpersonal differences in the *absolute* amount of utility are ignored. Therefore, not real individual utility is the basis of the analysis, but rather a normatively attached value of health, i.e. one equal for all persons. This may be considered inconsistent with for instance the model of Garber and Phelps (1997) since productive possibilities and income now may not determine the value of the QALY which is set equal for everyone. Since it cannot be denied that certain persons or groups in society or even in a certain health program *may* have higher utility levels than others, utility maximization is abandoned. In addition, both interpersonal differences within 'the health domain' and societal preferences of improving one person's health over that of another person are not considered in such an approach.

The difficulties in interpersonal comparison of QALYs by no means diminishes the value of investigating individual trade-offs concerning health, for example with standard gamble techniques. Still, it is most useful, not in the least for policymakers, to see how individuals come to decisions on health state changes and to find good approximations of individual's preferences for

[9] See Williams (1993) for a discussion on using average scores

health. It is the direct comparison of outcomes between persons and groups that should be considered with caution and we should especially be careful about how we interpret and present the results. At this point, the concept of QALYs and therewith the concept of cost-effectiveness analysis may better be categorized as an imperfect measure of health related quality of life, especially in the context of societal decision-making, but may be conveniently used as an aid to decision-making. This health outcome may very well indicate the 'health possibilities' for reaching the optimal amount of utilities, but should not be mistaken for cardinal, interpersonally comparable utilities.

In short, at this point completely embedding QALYs as utilities and therefore CEA in welfare economic theory seems impossible for programs in which interpersonal utility comparisons are necessary. This makes the criticism that the DMA is not fully embedded in welfare economic theory rather pointless. Currently, there appears to be no complete economic foundation of cost-effectiveness analysis that is applicable to the practical use of CEA.

Equity

Society and policymakers may incorporate equity concerns in decisions on the allocation of health care resources. One consequence in many western European countries is the existence of public health insurance, funded by obligatory contributions which are at least partly based on income. This should ensure equal access to basic health care for all members of a particular country. Within the health care system, the distribution of health benefits or health care is mostly related to the need of a patient and not to ability to pay. Indeed, as ability to pay therefore is often rejected as a method of distributing health care, the use of theoretical models in which an individual trades off health and wealth may be considered less appropriate to reflect real societal decision-making.

Van Doorslaer et al. (1993) have investigated policy statements of several countries on the subject of equitable finance and delivery of health care. Most statements indicate that the financing of the health care sector should be on the basis of ability to pay and should be unrelated to use of health care services. For instance, for Denmark Van Doorslaer and colleagues quote:

> *'Expenses are to be financed in the same way as expenses for other public services are financed, that is by means of taxes and duties which are adjusted to each individual's ability to pay.'*

The delivery of health care services should be based on need and access to the health care services should be ensured for all. Van Doorslaer et al. quote the policy statement of the United Kingdom:

> *'The government ... wants to ensure that in the future every man, woman and child can rely on getting ... the best medical and other facilities available; that their getting them shall not depend on whether they can pay for them or any other factor irrelevant to real need.'*

These considerations clearly indicate the importance of equity considerations in the allocation of health care services and the financing of those services. They also suggest that modeling of individually optimal solutions, in which the individual has to pay for his or her own consumption (trading off health and wealth), is not in line with these policy statements, nor with the health care systems as implemented in many countries. On a societal level a trade-off between health (care) and wealth must also be made, although this trade-off is not directly related to individual income. The result of this trade-off may be seen as captured in the health care budget, which is supplemented in decision-making from a societal perspective with information on other resource changes and budgets.

Also, in economic terms, distributing health care based on need implies that the initial distribution of wealth does not determine the distribution of health care and the latter is therefore implicitly rejected as a way of distributing health care goods. Of course, this is a fundamental difference from the Pareto-viewpoint, in which the initial distribution of wealth is not normatively appraised. Also, the redistributive aspects of taking money from the rich and healthy and giving health care to the poor and unhealthy may be interpreted as a violation of the Pareto-criterion, because the rich and healthy may be worse off because of this redistribution.

The strong conviction that health care should be distributed according to need has led to much discussion in the Netherlands on the subject of waiting lists. It can be argued that priority care for specific groups (employees) will lead to a reduction in waiting time for all patients when waiting time is reduced by extending normal capacity. If the costs of increasing effective health care capacity are paid by employers who benefit from reduced production losses, such a solution may be considered both efficient and

equitable, using the Pareto-criterion and the justice theory of Rawls (1971) (Brouwer and Schut, 1998). However, as this solution discriminates within the health care system on other grounds than medical need, it has been considered unacceptable and inequitable by policymakers and many social parties in the Netherlands, despite the possible reduction in waiting time for all. This further demonstrates the strong influence of ethics in medical policy and decision-making and how it may overrule efficiency gains.

The trade-off between equity and efficiency in health care is important, both on an aggregated level (e.g., Van Doorslaer et al., 1997) and for specific health care programs, for example for cervical cancer screening diminishing socio-economic health differences (Koopmanschap et al., 1990). Both welfarism and the decision-maker's approach have difficulties in fully incorporating equity considerations explicitly. Hurley (1998) even considers both welfarism and extra-welfarism to have a 'rather limited scope for accomodating equity concerns in the analysis.' One may feel that the scope for considering equity is even more limited in welfarism in which the Pareto-criterion and the initial distribution of wealth may be viewed as restricting the possibilities for redistributing health effects considerably (e.g. Rice, 1998). Also within welfarism, where the emphasis is on utility, it is again necessary to compare utility between persons and groups in order to assess which group should receive more (or less) weight (Bleichrodt, 1997). Moreover, if the weighting-process is not reflecting individual utility including altruistic preferences, some may consider weighting health effects as having no foundation in welfare economics (Johannesson, 1999).

If willingness to pay is used to value health effects, health care programs for people with higher socio-economic status would show higher benefits, ceteris paribus. In the case of mortality prevention, using a monetary value of a life saved may be considered equity neutral using an average value for all citizens, but this can hardly be reconciled with utility maximization principles underlying welfarism. A differentiation of the value by participation in paid work (or income) or by utility level again may have adverse equity implications. The US Panel on Cost-Effectiveness also encountered the difficulties of incorporating equity implications in an evaluation from working with welfarist models and principles. Weinstein and Manning (1997) for instance mention some areas of discussion within the US Panel and state:

'In each of these cases, there were equity implications of the economic welfare view that they were unwilling to accept.'

Weighting results, i.e. health effects, for different persons or groups is sometimes advocated as a means to incorporate equity considerations into the analysis. An important example hereof is the fair innings principle used by Williams (e.g. 1997), but others have also investigated the equity weights one might attach to health gains in different groups and persons (e.g. Johannesson and Gerdtham, 1996; Dolan, 1998).

In the recent discussion in the *Journal of Health Economics* between Dolan (1999) and Johannesson (1999) again a strict welfarist approach to societal decision-making versus a broader opinion about the nature of concepts of fairness and how these may 'shape the social welfare function' is encountered. Johannesson essentially takes a classical utilitarian social welfarist viewpoint, i.e., that social welfare is a sum of individual welfare. The additional element compared with the definition of welfarism provided by Culyer is that individuals are now allowed to be affected by the utility of others. On the other hand, Dolan asserts: '... I hypothesise that people may have preferences (for example, about social justice or fairness) that are extrinsic to individual utility.' By claiming that fairness may be something more than merely being affected by the utility levels of others in society, or *extrinsic* to individual utility as mentioned by Dolan, we move beyond strict welfarism. Moreover, following an approach such as that of Johannesson, one might feel to be dealing only with efficiency.[10] Although one may therefore agree with Johannesson that Dolan's approach of determining equity weights for gained QALYs has no theoretical foundation in welfare economics, this may sooner be the direct result of the narrow definition of social welfare used within strict welfare economics than anything else.

In the social welfare function there may be different ways of assigning weights to QALYs gained in different groups in society or to their relative opportunity of gaining QALYs (see for instance Bleichrodt, 1997). It may, however, in both views prove difficult to make all relevant weights and considerations explicit. The equity weights may differ per program, as these may be aimed at different groups in society. Furthermore, it involves not only distribution of health effects *within* one program, but also *over* different programs. In that

[10] Dolan (1999) points out that 'altruism is an efficiency consideration'

context, besides age *a priori* quality adjusted life expectancy of recipients of health effects from a certain program may influence equity weights as well (Waugh and Scott, 1998; Brouwer and Van Hout, 1998). Obviously, these different rationales for adjustments may interact and therefore it may prove difficult to formulate one strict decision rule. As earlier mentioned, not all equity concerns may be captured easily in utility terms or have a focus on utility. If the health care sector is aiming at producing *health* rather than utility, relevant equity considerations may also focus on health rather than on utility. Moreover, some equity principles may prove to be irreconcilable with the Pareto-criterion. In this respect the DMA provide analysts with more freedom to incorporate societal equity considerations in the analysis and societal preferences for health gains in certain groups that cannot be traced back to altruism.

Time costs and productivity costs

There has been a lot of debate concerning time costs and productivity costs related to absence from paid work in economic evaluations of health care. An interesting discussion in the current context is that between advocates of the human capital approach and those of the friction cost method. The former is sometimes praised for being consistent with neo-classical economic theory (e.g., Johannesson and Karlsson, 1997; Liljas, 1998). The latter however is seen as reflecting real world possibilities for replacing ill and disabled persons (e.g. Koopmanschap et al., 1995; Koopmanschap et al., 1997). Welfarists may consider the human capital approach as being superior for its foundation in welfare economics, but adherents to a more DMA-like approach may consider the standard neo-classical models as being inappropriate to reflect economic reality, especially from a societal perspective.

The US Panel on Cost-Effectiveness in Health and Medicine (Gold et al, 1996) also discussed this issue and advocated a new approach to incorporate productivity costs in an economic evaluation, incorporating productivity losses through income changes in terms of health benefits rather than in terms of costs. Although the US Panel begins by advocating a broad societal perspective in which CEA is an aid to decision-making rather than a full prescription for social choice (Weinstein and Manning, 1997), a view quite in line with the DMA, a quite restrictive individualistic welfare economic model, i.e. that of Garber and Phelps (1997), is used as a justification for their recommendation on productivity costs (Gold et al., 1996). Therefore, the societal DMA is not followed consistently, but is sometimes substituted by

individualistic welfarism, which may lead to conclusions that are difficult to reconcile with the general DMA perspective on CEA. Garber and Phelps show that in their model incorporation of production losses (which are equivalent to income changes) as either effects or costs will lead to consistent results. The US Panel subsequently recommends production losses to be measured in terms of QALYs, through consideration of income by respondents. To express productivity costs in monetary terms seems to be viewed by the US Panel as a movement towards CBA (Gold et al., 1996, page 182), and since income depends partially on health income changes may be seen as health effects.

The US Panel's rejection of a monetary incorporation of production losses sometimes seems related to the question of whether productivity costs express the value of life. In relation to mortality costs for instance, the Panel uses the term 'monetary value of life' when discussing the calculation of time costs related to mortality. Although the Panel rejects capturing these costs in monetary terms the terminology used is somewhat confusing. This confusion might be brought on by simple individualistic utility models, such as the one used by Garber and Phelps, in which the distinction between the *value of productivity* and the *value of life* is thin or even nonexistent. However, productivity costs due to morbidity or mortality are not necessarily equal to the value of life, but merely to the value of the *production* lost. Life has an intrinsic value (perhaps even more so on a societal level) which goes far beyond any productive capacity. Also, double-counting of productivity costs, as cautioned by the US Panel, will not occur when productivity costs are calculated in monetary terms and health states are assessed without consideration of income (e.g., Brouwer et al., 1997a; Brouwer et al., 1998). Savings in terms of reduced productivity costs may have equity implications for the results of an analysis. Inclusion of production losses and gains in economic evaluation may favour health care programs directed at employed persons over those affecting unemployed persons, since an equal reduction of illness will save more productivity costs in the former group. This situation could be easily avoided by excluding productivity costs from economic evaluations. However, this would ignore the fact that production losses influence the scarcity of resources and hence decrease the wealth of society. To give a full picture of indirect costs one should also value lost production related to unpaid work, diminishing adverse equity consequences. Furthermore, it would be advisable to report direct and productivity costs separately and to show the possible equity implications of including indirect

costs. It is the responsibility of decision-makers to decide on the relative weight that they want to attach to the equity considerations, separate from the relative efficiency of interventions.

Informal care

For many chronic and lethal diseases the role of informal care is substantial and probably growing, as a consequence of the aging population. Economists agree on the fact that if informal care is a non negligible part of a specific health care program, it should be incorporated into economic evaluation as it involves the use of scarce resources. Regarding the valuation of informal care health economists are often somewhat vague. However, it appears that welfarists prefer to value informal care monetarily, using individual estimates of the opportunity costs of time (paid work, unpaid work and leisure). Mostly, this is approximated by the wage rate, assuming that the labour market is capable of equalizing the marginal value of a unit of leisure time or unpaid work and the market wage rate. If informal caregivers do not have paid work, a general age and sex specific wage rate is often used as an approximation.

Following the DMA, some analysts would simply present the amount of informal care, without a valuation, whereas others would use the actual wage rate, a reservation wage rate for those currently not working or the market wage for household work in case of unpaid work. A third possibility is suggested by Brouwer et al. (1999). They propose to value informal care not only monetarily. In valuing the time input they suggest the friction cost method for paid work and a market price for household services as a shadow price for unpaid work. The positive or negative process utility related to the caregiving process as such should be valued by quality of life measurement. Quality of life of informal caregivers should be a separate result in economic evaluation, not to be summed simply with patients' quality of life. For leisure time sacrificed Brouwer et al. do not advocate a monetary value, since the end of leisure time is not production but enjoying the pleasant things in life. Hence, they propose to value lost/gained leisure time of informal caregivers in terms of quality of life. In this way, also intangible effects like fatigue and distress from seeing one's partner suffer may be captured in terms of quality of life. Of course, following such a proposition again makes the result of a CEA more extensive than a simple C/E-ratio, as it is not without difficulties (if at all feasible) to combine both informal caregiver's and patient's quality of life. On the other hand, taking a societal viewpoint in economic evaluation does encompass considering all costs and health effects regardless of who

bears these costs or who experiences those health effects. To confine this broad definition to patients only may be considered inappropriate.

Discounting

Most economists seem to agree that future costs related to a health care program should be aggregated over time using a discount rate which reflects the average return on relatively low risk investments such as long term government bonds (Drummond et al. 1997). Regarding time preference for health effects the views are more divergent. Some welfarists feel that individualistic preferences for current versus future health should be the basis for discounting effects. The available research on elicited time preference shows a wide variation in discount rates, but in general these rates turn out to be quite high: for example 19-24% per year according to Cairns (1994). Using these time preferences in economic evaluation would imply that health care programs that have some delay in producing health effects turn out to be quite unfavourable. This effect may be somewhat less when using a hyperbolic discount function, discounting the near future more than the far future (e.g. Cairns and Van der Pol, 1997).

Within DMA, analysts tend more towards a discount rate for health effects set by the decision-maker, reflecting societal time preference for health effects, probably correcting for myopic individual attitudes in a somewhat paternalistic way (Krahn and Gafni, 1993). Many analysts seem to choose for a 5% discount rate (equal to costs), because they are convinced by the Weinstein and Stason (1977) argument that a lower discount rate for health effects would lead to implausible results[11]. Some may be convinced by Keeler and Cretin (1983) that using different discount rates for costs and effects leads to infinite postponing of programs.[12] However, one may question whether this mathematically correct doom-scenario is relevant in the policy context of fixed health care budgets which have to be assigned to different programs each year. As an unfortunate consequence, analysts do not feel the need to underpin the choice for a specific discount rate for health effects. However, we expect that debate on this issue will continue in the nearby (undiscounted!) future.

If and how costs and health effects which accrue to the next generation should be discounted is a related issue. It may be the case that our concerns

[11] For a discussion on the validity of this argument see van Hout, 1998

[12] See also chapter 9

for the health of our children and grandchildren are not adequately reflected by standard discounting. Notwithstanding this, Johannesson and Johansson (1996) ask people to elicit individual time preferences regarding health for future generations. Using binary choices between saving lives in the current generation versus saving lives in future generations, they estimate an annual discount rate of 25% for a 20 year horizon, 12% for a 50 year horizon and 8% for a 100 year horizon. Gyrd Hansen and Søgaard (1998) have recently tried to reconcile the individual and the societal viewpoint in discounting, proposing a two-stage discount model, as previously discussed by Libscomb (1989). Gyrd Hansen and Søgaard propose a measure of social interpersonal time preference and individual time preference for discounting changes in life expectancy; the social intertemporal time preference is then used for discounting the health benefits from the moment of risk reduction to present time, while the individual time preference is used from the start of treatment for all years in which health effects are gained. This combination may also be considered related to the view that individuals may have a discount rate that they 'use' in their private sphere, while they want policymakers to use another discount rate for the decisions in the societal domain.

Here, again the general view on CEA may determine in which direction a solution to the problem is sought. Where welfarists may strive to improve the determination of the discount functions in individualistic behaviour, DMA-adherents may leave the choice up to decision-makers or try to find some more normative decision rules reflecting society's time preference or which may be endorsed by society (but may be 'paternalistically' divergent from individual discount behaviour). In the end, decision-makers have to decide between different opinions on this matter and on equity between generations.

2.4 Discussion, conclusion and recommendation

After having discussed some differences in methodological choices between welfarism and the DMA, in this section we want to address some related questions and provide some conclusions and recommendations.

Multi- or monodisciplinary research?
Should economic evaluation be a task predominantly for health economists or preferably carried out by a multidisciplinary research team? In strict welfarist research one may feel that economists have a rather dominant position in

research, leaving only restricted room for psychologists and physicians in describing health effects. In the DMA the health economist may be seen to inherently receive a more modest position. Health effects are not monetarised or seen as strict utility and input from other scientific fields in assessing these health effects are indispensable. For instance the burden of informal care can be modeled taking non-economic and non-utility arguments into account. Consequently, economic evaluation may get a somewhat multicriteria analysis-like character, because not all important aspects of health care programs can be fitted into a C/E-ratio.

The role of the decision-maker

It may be argued that DMA puts the decision-maker in a more difficult position, as one does not dispose of a clear-cut net benefit figure as in CBA or a simple CE-ratio involving utilities in CEA. However, one could argue that providing the decision-maker with a systematic overview of societal costs, health effects, together with equity implications, quality of life of informal care and other important considerations, would provide the decision-maker with the opportunity to give explicit weights to the relevant criteria. This may enhance the transparency of the decision-making process. It should be admitted that this requires the decision-maker to have better insight in the strengths and weaknesses of economic evaluation, its inherent normative value judgements and the limitations of its use. In our opinion this approach has the advantage that the primary responsibility for health care decision-making is put where it belongs: in the political arena, with a minimum of value judgements (if so then explicitly mentioned) made by the researchers themselves.

Conclusions and recommendations

First of all, perhaps we should acknowledge that welfarists and DMA-adherents have different aims with CEA. Whereas adherents of the DMA want to make the process of decision-making in the health care sector more uniform and rational, without focusing on strict compliance with welfare economics, welfarists may consider the latter to be more important. This implies a difference in focus and, in the end, a choice between either side probably depends on the purpose of the researcher. We do feel that for practical purposes at this point the DMA may be more appropriate, because no full applicable welfare economic framework for decision-making exists and because the goal of policymakers in health care may be inherently different from maximizing utility.

Secondly, depending on the general viewpoint taken on CEA, different methodological choices will prevail. It is good to keep this in mind and where possible try to specify the viewpoint chosen. Welfarists may wish to reconsider the validity of interpreting QALYs as utilities *in a setting of societal decision-making.* If CBA is chosen as the mode of analysis the monetarisation of health benefits (through willingness to pay) has to be investigated for equity implications. In general, it should be examined whether using a societal perspective in economic evaluation can be reconciled (also in practice) with the focus on (individual) utility in welfarism and basing methodological recommendations on individualistic models. Analysts following the DMA should recognize that the results of CEA/CUA contain much more than a cost effectiveness-ratio, and also may have the form of a multicriteria analysis. If the aim is to take a societal perspective, this starting point should be pursued in all aspects of the analysis, especially those concerning time and productivity costs, quality of life and discounting. There should be closer co-operation between researchers and decision-makers, in order to outline the reporting format of economic evaluations, paying attention to explicating the value judgements underlying the results.

Finally, in this chapter we may have overstated the differences between the welfarist view and the DMA in order to clarify our arguments. Of course, not all differences in opinion are consequences of these divergent views on CEA. In addition, the aim of both lines of thought may not be that different, but it may be more a difference in the way in which matters are operationalized and interpreted. However, we do hope that this overview of concepts used sheds more light on the contents and origins of current controversies and disputes between health economists involved in economic evaluation of health care. In addition, we hope that this chapter contributes to constructive discussions on some of the issues raised and may help to attain a more uniform methodology in CEA, so that our choices in health care may be increasingly based on a sound decision-making procedure and reasoning. As Sen (1995) writes

'*While Aristotle agreed with Agathon that even God could not change the past, he did think that the future was ours to make — by basing our choices on reasoning.*'

Acknowledgements

We are grateful to Han Bleichrodt, Ken Redekop, Frans Rutten, Fern Terris-Prestholt and two anonymous referees for their useful comments on an earlier version of the manuscript.

3. Patient time in cost-effectiveness analysis

Summary

The time invested by patients in treatment and rehabilitation should be valued in a cost-effectiveness analysis. The U.S. Panel on Cost-Effectiveness in Health and Medicine gives recommendations on incorporation of time costs that lead to a misrepresentation of the true societal costs. This chapter provides alternative recommendations for incorporating costs of time in CEA. Following an opportunity costs approach time is separated into three parts, each with its own valuation methods: time spent on paid work, unpaid work and leisure.

3.1 Introduction

Time is an important resource input in health care. Not only do doctors and nurses spend time treating and caring for patients, but patients themselves also invest a considerable amount of time in assessment, treatment and rehabilitation. The time invested by patients can be seen as a direct cost (Drummond et al., 1995), but appears rather difficult to value. As yet there is no clear consensus on one or the other method. The relevance of carefully valuing this type of resource-use in cost-effectiveness analysis differs from one study to another, since the amount of patient time spent ill, in treatment and in rehabilitation vary substantially across different diseases and interventions.

The main question to be answered in this chapter, is how to value patients' time from a societal perspective, when they fall ill, are being treated or are recuperating. The opportunity cost method is often used to value time from the patient's perspective. In the recently published guidelines of the U.S. Panel on Cost-Effectiveness in Health and Medicine (Gold et al., 1996) attention is paid to the problem of time costs of patients. The U.S. Panel has taken on the difficult task of finding a way to value time of patients in a cost-effectiveness

analysis. However, other recommendations for incorporating these time costs in CEA may be put forward, as will be explained below.

In section 3.2 the relevance of the concept of patient time as input is discussed. Then sections 3.3, 3.4 and 3.5 discuss patient time otherwise invested in paid work, unpaid work and leisure respectively. Each of these sections discusses the opinion of the U.S. Panel, presents our comments and provides alternative suggestions for incorporation of patient time costs in a cost-effectiveness analysis. Section 3.6 considers time of children, elderly and disabled persons. Section 3.7 concludes this chapter.

3.2 Patient time as input from a societal perspective

Patient time is recognized by the Panel as a distinct resource in cost-effectiveness analysis. Patients must give up alternative uses of their time because of treatment or just being ill. Although there is no direct compensation for this time, it is not free in the economic sense of being without value. The activity in which one would normally have been engaged during the time of illness and/or seeking, undergoing and recovering from treatment is considered the best alternative use of time and the value of that activity is equal to the opportunity cost.

The Panel makes a distinction between patient time that is invested in the treatment (time costs) and other time (morbidity costs) and suggests different approaches for their valuation. The selected approach is dependent on 'somewhat arbitrary' (Gold et al., 1996, page 182) decisions on what time can still be seen as investment and what time cannot (for example, waiting for treatment, rehabilitation, etc). Morbidity costs, the costs associated with the time that is not invested in the treatment, like time spent ill or recuperating, can be seen as indirect non-medical costs, or 'productivity costs', which the Panel suggests as a new term for these costs. These costs are normally estimated with the friction cost method or the human capital method. The Panel defines these costs as the

'*costs associated with lost or impaired ability to work or to engage in leisure activities due to morbidity, such as time for recuperation and convalescence*' (Gold et al., 1996, page 181).

The Panel contends that these costs should be incorporated in the denominator of the cost-effectiveness ratio. The main reason for this recommendation is that respondents to questionnaires used in QALY measurement will take into account lost earnings due to impaired ability to work if these instruments used 'are silent concerning the consideration of lost income' (Weinstein et al., 1996). The Panel makes an exception only for instruments that explicitly exclude the consideration of lost income and excludes these instruments from their 'reference case analysis'. It is questionable whether the recommendation of the Panel on this point will lead to good estimations of productivity costs from a societal perspective (Brouwer et al., 1997a; Brouwer et al., 1997b; Weinstein et al., 1997).

Since the main concerns about the recommendations related to morbidity costs are discussed in chapter 4 and 5, the focus in this chapter is on the valuation of patient time costs, the costs associated with time invested in the treatment. However, it is stressed here, that the presented recommendations on the incorporation of patient time may be used for all patient time in the analysis, since it is irrelevant in our opinion whether the patient's time is invested in treatment. From the societal point of view it only matters what would have been produced, or what pleasant activities a person would have engaged in, if the illness could have been prevented, shortened or limited. The resource involved is time of the patient, the value of which should be determined regardless of how it is used in case of illness. How it would have been spent otherwise (following an opportunity cost method) is the relevant question here.

Time spent in treatment is valued differently from other time in the recommendations of the Panel. We believe that it is useful and important to value all time of patients in one uniform way, because it may be difficult to disentangle treatment and nontreatment time (e.g. in case if kidney dialysis, where treatment becomes part of 'normal' life almost). In addition, some types of treatment, such as chemotherapy or radiotherapy, may have additional impact on quality of life. These effects on quality of life, although perhaps temporary, may have a substantial impact on the patient and his or her environment, and also on treatment compliance. The Panel argues that these quality of life effects should be additionally measured in terms of quality of life (Gold et al., 1996). Our aim is to provide one uniform set of recommendations that guides the valuation of patient time. Quality of life then is an integral part of the valuation of patient time.

Definitions
Changes in time use will have two impacts. First, productivity costs may occur due to reduced ability to perform paid or unpaid work. These productivity costs may be defined as follows:

Productivity costs are costs associated with production loss and replacement due to illness, disability and death of productive persons, both paid and unpaid.

Besides productivity costs there may also be direct impacts on quality of life. For instance, impaired role functioning or the lost ability to engage in leisure activities may affect quality of life directly. These quality of life aspects of the changed time use should also be incorporated into the analysis. Together with productivity costs these costs are the total impact of changed time use. These 'time costs' may be defined as follows:

The patient's 'time costs' are equivalent to productivity costs plus the impact of changes in time use on health-related quality of life, as a result of illness and disability.

3.3 Paid work

When considering a patient who is absent from a paid job, the wage rate is seen as the relevant opportunity costs by the U.S. Panel. If the illness had not occurred, the patient could and probably would have worked for a wage and therefore the opportunity costs per hour are equal to the wage rate per hour. Thus, the value of time is valued by multiplying the wage rate per hour (day) with the relevant number of hours (or days). This is, in fact, similar to the human capital method, although after undergoing treatment the change in income due to changed productivity is recommended to be incorporated in terms of quality of life by the Panel, as mentioned. However, as critics of the human capital method have indicated (e.g. Koopmanschap et al., 1996), this valuation method may not hold from a societal viewpoint, because of the possibility of replacement. Since cost-effectiveness analysis is to be performed from a societal perspective all significant impacts should be incorporated in the analysis. With respect to disability to work two parties are especially important: the person falling ill and the person replacing him or her, here assumed to be a formerly unemployed person. The first will incur a labor-income decrease, while the latter will incur an income increase. The

Panel clearly states that the societal perspective 'is the only perspective that never counts as a gain what is really someone else's loss' (Gold et al., 1996, page 7). Similarly, we should also not count as a loss the lost wage of the patient, what is really someone else's gain, i.e. the gained wage of the person replacing the patient. For instance, if a patient is absent from a paid job to receive intensive treatment and rehabilitates during one year, it does not necessarily mean that society incurs production losses during the total period of one year. In most economies a person may be replaced after some time, possibly by a formerly unemployed person. The costs for society in the case of replacement after a certain period are the lost value added because of lost productivity caused by absence of the patient and the replacement costs. The production level may return to its original level after replacement, but the selection and hiring processes take some time. During this *friction period* society incurs productivity costs that are measured by the friction cost method. The human capital method measures the potential rather than the actual production losses. The outcomes using the friction cost method are therefore lower than with traditional human capital estimations (Koopmanschap et al., 1995). This is illustrated in Table 3.1.

Cost category	Human capital method	Friction cost method
Absence from work	23.8	9.2
Disability	49.1	0.15
Mortality	8.0	0.15
Total indirect costs	80.9 (18%)	9.5 (2.1%)

Table 3.1: Indirect costs of disease in the Netherlands for 1988 in billions of Dutch guilders (as percentage of NNI).
(Source: Koopmanschap, 1994)

Since the friction period starts from the incidence of absence from work, the uniform treatment of all patient time in cost-effectiveness analysis facilitates the use of the friction cost method, because there is no distinction between time in treatment and other time. For a more fundamental discussion of the friction cost method we refer to the recent discussion on this subject (Koopmanschap et al., 1995; Johannesson and Karlsson, 1997; Koopmanschap et al., 1997).

Above mentioned productivity costs are not the only impact from a societal perspective. Not being able to perform a paid job anymore is also associated

with a reduced quality of life because of impaired role functioning. The reduction in quality of life due to the impaired *ability* to function in society is clearly health-related. This *inability* of functioning in a paid job should therefore be measured as reduced health-related quality of life of the patient and be incorporated in the denominator of the cost-effectiveness ratio.

These arguments lead to the following two recommendations.

Recommendation 1: The time spent less on paid work by the patient can best be valued with the friction cost method. These are productivity costs.

Recommendation 2: The reduced health-related quality of life of patients due to impaired ability to function in society can best be measured by means of a general and validated quality of life questionnaire.

3.4 Unpaid work

Many persons do not have a paid job, yet are engaged in productive activities at home, like preparing meals and taking care of children. The U.S. Panel recommends that

'*when a person is of working age but does not work for pay ... one option is to use the hourly wage rate of individuals with the similar characteristics ... who do work for pay.*' (Gold et al., 1996, page 202)

The Panel contends that since the person chooses to work at home rather than for a wage, the wage rate is a lower bound estimate that 'may be close enough to the real opportunity costs of time'. The characteristics of the patient are used to establish the wage rate that a similar person with a paid job receives, and this wage rate is used as a proxy for the costs of lost home production.

One objection to taking representative wages that one could have earned on the labor market as a proxy for opportunity costs is the difficulty of capturing monetarily the real opportunity costs of home production. The choice between a paid job or working at home (supposing there is such a choice in case of considerable unemployment) is not merely a trade-off between the value (in money) of the home-production and the wage one would earn as

employee. The direct utility from caring for one's children, making one's home more comfortable, etc. (process-utility), added to the value of the home production itself (outcome-utility or product-utility) form the total gain from working at home. The Panel acknowledges that 'if someone enjoys helping others or caring for children ... the wage rate used should be revised appropriately' (page 202). However, this process-utility from caring for one's children, etc., is not believed to be easily expressed in monetary terms, and therefore the appropriate revision of wage rates may prove to be extremely difficult, if not impossible. Taking the unadjusted wage as proxy for the value of the home production including this process utility may not be that close to the real opportunity costs of time. Especially for people engaged in home-production, who would earn very little in a paid job (for instance because of a low educational level) it is not clear why caring for children should be valued relatively low according to their potential wage rate, while for others, who may have greater opportunities in a paid job, this should be valued higher.

Furthermore, looking at outcome, it is hard to maintain that, for instance a home-cooked meal prepared by a health economist is itself worth more than that of a person with no education. Wages cannot always reflect one's ability to perform activities at home in a valid way. Taking the wage of a professional housekeeper as an approximation of the value added of home production as such seems the most appealing method to measure the value added of a person engaged in home-production.

In sum, the outcome utility of home-production can best be valued equal to the wage rate of a professional housekeeper. This monetary value of home-production as such is, as mentioned, not the only quantity lost when someone can no longer engage in home-production anymore. The reduction in quality of life from not being able to engage in these activities, the lost process-utility, may be measured meaningfully in terms of quality of life, since it does not necessarily correspond well with potential wage rates. This impact on the quality of life of a patient may be seen as "home role-functioning" and thus be treated like other forms of role functioning, i.e. valuation through quality of life measurement. To some extent this decrease in quality of life is already measured in quality of life questionnaires. For instance, EuroQol/EQ-5D and SF 36 ask about the effect of illness on performing housework. This implies that the approach of the Panel may lead to double-counting, when additional quality of life measurement is required.

Recommendation 3: The value of home-production can best be estimated by taking the wage rate of a professional housekeeper. These are productivity costs.

Recommendation 4: The reduction in quality of life from being less able or unable to engage in home production activities that contribute positively to one's quality of life can best be measured in terms of quality of life.

3.5 Leisure time

Leisure time is a third major time use. The Panel states that the wage rate of a person with similar characteristics working for a pay 'generally does not reflect the value of time for persons primarily engaged in leisure'. Therefore,

'for individuals not engaged in compensated employment wages used as proxies must be adjusted to reflect the full opportunity cost of time.' (Gold et al., 1996, page 210)

Similar approaches to that of the Panel can be found in transport appraisals, where travel time must be valued (Smith and Wright, 1994). Also, from a neo-classical viewpoint, the wage rate in some cases may be used as a proxy for costs of forgone leisure time (Posnett and Jan, 1996). It should be noted that the Panel requests more research on how to implement this approach rather than pointing out how to place a monetary value on leisure time.

In our opinion, leisure time may best be treated as time in which one can do the things that make life valuable: sporting, hobbies, socializing, etc. The loss of leisure time can be captured most meaningfully by measuring the reduction of quality of life from lost leisure time, now spent ill, in treatment or recuperating. The lost ability to engage in leisure activities can be captured in terms of health-related quality of life and thus be incorporated in the denominator of the C/E ratio. Johannesson recently expressed a similar opinion (Johannesson, 1997). In this way, lost leisure time does not have to be calculated in monetary terms, which obviously is not easily done. Furthermore, EuroQol/EQ-5D and SF-36 explicitly ask about problems in performing leisure activities or social activities. To count the loss of leisure in monetary terms in the numerator as well may lead to a double count, which should be avoided. However, in additional quality of life measurement, as proposed by the Panel, it will be difficult to avoid these problems.

Recommendation 5: Patients' lost leisure time can best be valued through quality of life and hence should not be valued in monetary terms.

3.6 Time of children, elderly and disabled persons

The valuation of time of children, elderly and people that are unable to work proves to be just as difficult as the valuation of leisure time. Since people in these groups normally are not engaged in paid work, an (adjusted) wage rate again is not easily found. The Panel encourages research on this subject and provides some suggestions for monetary valuation of the time of these three groups. The Panel argues that for teenagers it may be possible to provisionally base 'the wage rate on teens in the labor force and adjusting as necessary for the selection bias of using observed market wages for teens not in the labor force' (Gold et al., 1996, page 202). The approach of the Panel is aimed at attaching monetary values to all types of time use in order to capture the full opportunity costs of time of children, elderly and people who are unable to work.

In our opinion, the time of children and people who are permanently disabled is difficult to capture in a monetary value, especially when wages should provide a basis for valuation. Time of the elderly, who no longer work for a pay, may be valued by a combination of valuation methods as described in sections 3.4 and 3.5, since they may still be engaged in leisure and home production.

Children's time cannot be valued meaningfully based on their current productive activities, since they normally do not have a productive task in developed societies. The provisional solution of the Panel to value time of teenagers according to the wage rate of working teens may not be valid since this approach does not capture the full opportunity costs of children's time. Such a method would imply that time of children would be worth close to nothing in comparison to adults' time. In our opinion, children's time can best be valued in terms of quality of life. The same holds for disabled persons. When these persons are normally (and significantly) engaged in unpaid work, this time might also be valued with the method as described in section 3.4.

A noteworthy aspect of children's time is education. If a child has to undergo treatment and has to postpone or even cancel his or her education, this may limit his or her future productive capacities. This aspect is hard to capture in quality of life measurement or to quantify monetarily, yet it may cause societal costs and should in principle be incorporated in the analysis.

Recommendation 6: The time of the elderly who do not work for a pay can best be valued through quality of life measurement, except for lost output from home production, which should be valued as described in recommendation 3.

Recommendation 7: The time of children and those unable to work can best be valued through quality of life measurement.

3.7 Discussion

A noteworthy aspect in this chapter is the separation of quality of life aspects related to time from the costs related to time. Together these two parts should measure the full impact of changes in time use. Because the recommendations in this chapter aim at making quality of life measurement and costs measurement complementary, it is necessary to carefully choose a quality of life instrument that indeed fully captures the quality of life implications of changes in the different types of time use, such as the EuroQol/EQ-5D. In this way, one ensures the incorporation of the total impact of changes in time use. For the operationalization of the cost-measurement, the practical guide for using the friction cost method may be useful (Koopmanschap and Rutten, 1996a).

A number of areas remain open for future research. One important research area is the process of replacement when someone engaged in home production activities is unable to continue normal activities. When a person engaged in home production activities is replaced by an unpaid person, such as a family member, this person must invest time to perform these home-production activities, causing opportunity costs such as lost leisure time. Exactly how home production is continued when a homemaker falls ill and at what costs should be investigated further.

Further research may also be aimed at finding or (further) constructing quality of life questionnaires and costing principles to be used in cost-effectiveness

analysis, that avoid double-counting and are sensitive enough to measure the impact of changed time use. Finally, the effect on future productive capacities of children who miss out on education due to illness and treatment should be investigated further.

4. Productivity costs measurement through quality of life?

Summary

This chapter comments on the recently published guidelines of the U.S. Panel on Cost-Effectiveness in Health and Medicine on incorporation of indirect non-medical costs, or productivity costs, in economic evaluations of health care. Traditionally the human capital or more recently the friction cost method is used to measure these costs. The Panel, however, recommends to incorporate productivity costs as health effects in the denominator of the C/E-ratio. This chapter argues that incorporation of productivity costs in cost-effectiveness analysis expressed as health effects is not correct. Only direct health related effects on quality of life that cannot be meaningfully monetarized should be considered as health effects. Furthermore, measuring productivity costs in terms of quality of life may lead to misrepresentation of these costs from a societal viewpoint. This misrepresentation occurs because of the existence of social security systems and private insurance compensating for income reductions from disease. Furthermore, the patient's viewpoint is useful for quality of life measurement, but not for measuring productivity costs from a societal perspective. Finally, alternative recommendations are formulated for incorporating societal productivity costs in economic evaluations of health care.

4.1 Introduction

In recent years there has been much debate about if and how indirect non-medical costs should be included in a cost-effectiveness analysis. The size of these costs may be considerable and since they constitute a cost to society they should be incorporated in an analysis taking the societal perspective. Although the inclusion of these costs is debated, the consensus on incorporating these costs seems to be growing. Canadian, French, German and British guidelines are in favor of including the indirect non-medical costs, as is the U.S. Panel in its recently published guidelines (Gold et al., 1996). How to value the indirect non-medical costs is more difficult. Traditionally

the human capital method was used in calculations, but more recently the friction cost method (Koopmanschap et al., 1995) has gained attention and is increasingly mentioned in guidelines as the proper method to use.

In this chapter first indirect non-medical costs are defined and described, after which the human capital method is compared briefly with the friction cost method. Then, in response to the recently published guidelines for economic evaluations of health care interventions by the Panel on Cost Effectiveness in Health and Medicine (Gold et al., 1996), the best way to incorporate *productivity costs* into cost-effectiveness analysis is discussed. Productivity costs is recommended by the U.S. Panel as an alternative term for indirect non-medical costs. We support this new term, which shows clearer what is incorporated in these costs, although the term 'costs of lost productivity' would be less of a *contradictio in terminis.*

The important question in this chapter is how to measure productivity costs from a societal point of view and how to incorporate these costs in cost-effectiveness analysis. Taking the societal viewpoint implies that all costs and benefits relevant to society should be taken into account, regardless of who bears those costs or who receives those benefits, in contrast to merely looking at costs and benefits from the perspective of a single actor. The main focus in this chapter will be on societal productivity costs in relation to paid labor, since measuring and valuing productivity costs related to unpaid labor is a different subject with very specific problems requiring special attention (see chapter 3).

4.2 Productivity costs: definitions and measurement

First a proper definition of productivity costs should be formulated. The U.S. Panel gives the following:

'... costs associated with lost or impaired ability to work or to engage in leisure activities due to morbidity and lost economic productivity due to death.' (Gold et al., 1996, page 181)

We propose another definition of productivity costs:

Costs associated with production loss and replacement costs due to illness, disability and death of productive persons, both paid and unpaid.

Both definitions allow for productivity costs to occur while working, but being less productive. This type of productivity costs is often not discussed, but may be quite relevant for some diseases such as migraine (Van Roijen et al., 1995a; Osterhaus et al., 1992).

Both definitions of productivity costs suggest that when people's productivity is affected by morbidity or mortality, society may incur costs. If a paid employee is ill and has to stay at home, his employer may experience a production loss. The value of this production loss is a cost to society. Replacement costs in case of longer absence spells (such as advertisement costs, costs of selection and costs of training of a new employee) should be seen as productivity costs as well.

The U.S. Panel includes the lost or impaired ability to engage in leisure activities in their definition of productivity costs. We would like to argue that these 'costs' affect quality of life and should be counted as health effects, and should not be seen as productivity costs, since engaging in leisure is consumption rather than production following the third person criterion as introduced by Reid (1934) .

Traditionally, the human capital method was used for measuring productivity costs. It is a rather simple and straightforward method in which lost production is measured using gross income on the individual level. Regardless of the length of the period of absence, the gross wage that a person would have earned during absence, is taken as an approximation of lost production. For example, if a person would suddenly die or become disabled at age 28 earning $50,000 a year at that moment, the human capital method counts his wage[13] of $50,000 times the number of years that he would have worked if he were still alive. Aggregation of these individual figures yields the productivity costs at the national level.

The argument for using the wage rate to value lost production is based on the neo-classical viewpoint that wage rates are the equivalent of the value of marginal productivity. Therefore, all earned income can be taken as the monetary value of all production lost. This also holds for an analysis from a societal point of view in the textbook situation under strict assumptions.

[13] Sometimes a prediction of average wage (since wages tend to increase during ones career) is used to make the lost income estimation more realistic .

A major textbook assumption is that of full employment. If the unemployment rate in an economy is higher than the frictional unemployment, then it will often be possible to replace a person (after a while). However, in a textbook analysis as described above replacement is impossible and hence production losses continue until the retirement age. The friction cost method, however, does allow for unemployment and other realistic circumstances to exist in its productivity costs calculations as it distinguishes between a friction period, in which productivity loss occurs, and a further period where a sick worker has been replaced.

From a societal point of view only the value of production that is lost until the return of the absentee or until someone replaces the absentee plus the possible replacement costs may be counted as productivity costs. If, for instance, a person becomes disabled and his employer finds replacement for him after 6 months by hiring a qualified unemployed person, from a societal point of view production then returns to its initial level and no additional productivity costs after replacement are incurred. The only period that productivity costs may occur is within the *friction period*, that is the period between the absenteeism and replacement. Within this period replacement costs are incurred and/or production losses are suffered. These friction costs represent the only costs from a societal point of view. The lost income of the disabled person is now being earned by the formerly unemployed person (if the wage rates are comparable) and there is no lost production or lost income from a societal viewpoint. From a societal point of view it makes no difference whether person A earns nothing and B does or vice versa. If there is a difference in the level of social benefits between person A (unemployed) and B (disabled), then this difference can cause some medium term macro-economic consequences which may be estimated through a macro-econometric model (Koopmanschap et al., 1995).

The friction cost method is to be preferred over the human capital method since it allows for disequilibria in economies, such as unemployment. The total indirect costs due to disease, disability and death for the Netherlands for 1988 with the human capital method was calculated at 80.9 billion guilders, whereas the friction cost method estimated these costs at 'only' 9.5 billion guilders (Koopmanschap et al., 1995), implying a difference in results of more than a factor 8. The importance of estimating the productivity costs realistically is also demonstrated by the fact that indirect costs may constitute

a large fraction of total costs e.g. in certain disease categories, such as migraine (Van Roijen et al., 1995b). Differences in unemployment rates between countries may influence the outcome of the friction cost method. These differences should be detected and presented and may possibly be corrected for in international comparative studies, just as for differences in direct costs, which are often considerable.

4.3 Recommendations of the U.S. Panel

In their recently published guidelines, the U.S. Panel also pays attention to the proper way of calculating productivity costs. Although some of its recommendations are very useful, the approach chosen by the U.S. Panel on how to incorporate productivity costs in CEA (i.e. in terms of health effects) is not correct in our opinion. In this section we will briefly discuss the recommendations of the U.S. Panel and in the next section we will provide our arguments for recommending a different approach.

The U.S. Panel recommends measuring most of the productivity costs through quality of life measurement in the *denominator* of the C/E ratio in terms of health effects, combined with extra *friction costs*, which they only seem to define as replacement costs, such as training costs and '... if substitute labor is never quite as productive as the labor it replaces and the difference is not fully captured by wage rates, then the discrepancy is a cost' (Gold et al., 1996, page 183).

These replacement costs are calculated in monetary terms and included in the *numerator* of the C/E ratio.

On the health effect side, the Panel distinguishes between two different types of impacts from reduced productivity: impaired role functioning and income reduction. The first has always been included on the effect-side, but both the human capital method and the friction cost method measure productivity costs related to production loss or income reduction in monetary terms. This implies that they should be incorporated fully in the costs of an intervention (in the numerator of the C/E ratio), thus the U.S. Panel proposes a rather new approach.

The U.S. Panel states that a choice between inclusion of the productivity costs in the numerator or the denominator should be made, because to count both quality of life-effects of decreased productivity and all costs normally measured with the friction cost or the human capital method would lead to a form of double counting which obviously should be avoided. The Panel goes on by stating its preference to include these costs in the denominator, in terms of health related effects. An important consideration in the U.S. Panel's recommendation is that the loss of income is assumed to be incorporated by the respondents in a quality of life measurement, when the instruments used do not explicitly exclude the consideration of lost income. Thus, people are assumed to assess quality of life as compared with what used to be normal, including the change in quality of life that originates from income losses due to being unable to work. Also, the U.S. Panel feels that including productivity costs expressed in terms of quality of life on the effect side, therefore in the denominator of the C/E ratio, '...conforms more closely to the principle of cost-effectiveness analysis - namely that 'effects' are included in a nonmonetarized form in the denominator' (Gold et al., 1996, page 181-182).

Quality of life of persons is believed to depend, in part, on role functioning, as mentioned. Persons tend to indicate a higher quality of life when they can function well in their social setting, in their job for instance. The U.S. Panel recommends that impacts on role functioning should be measured as health related quality of life-effects in a cost-effectiveness analysis and thus be incorporated in the *denominator*.

The impact of an income reduction on quality of life will be included in the *denominator* of the C/E ratio as well. This relationship between productivity, income and quality of life implicitly assumed in the reasoning of the U.S. Panel is presented in figure 4.1.

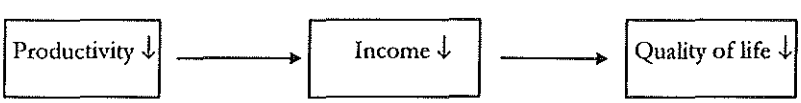

Figure 4.1: The relationship between productivity, income and quality of life.

Figure 4.1 shows that the relationship between productivity and quality of life runs partially through the patient's income. From this figure it becomes obvious that a productivity reduction causing a decrease in income in turn will

lead to a reduction of quality of life. The patients will base their estimated quality of life reduction on the loss of income rather than directly on the loss of productivity. Therefore, the link between productivity and income becomes crucial in the measurement of productivity costs as recommended by the Panel.

4.4 Productivity costs incorporated in quality of life measurement: some objections

The approach as recommended by the U.S. Panel does not lead to an accurate estimation of the productivity costs from a societal point of view. First, we will address the choice between numerator and denominator as the appropriate place to capture productivity costs. Then, the way the patient will value 'his productivity costs' in quality of life measurement, if asked to do so, is discussed and finally, the question how these individual quality of life measurements and costs as proposed by the U.S. Panel relate to the societal productivity costs is discussed.

Numerator versus denominator

The U.S. Panel holds the opinion that including the productivity costs in quality of life and thus in the denominator of the C/E ratio is acceptable, since theoretically it does not matter whether one incorporates all relevant impacts as costs or effects as long as all impacts are captured in the analysis and double counting is avoided. However, if we look at the reason why cost-effectiveness analysis has become such an important instrument, it is clear that we should not incorporate quantities that can be valued monetarily into the health effect-side of the analysis too easily.

It is obvious that the health effects became a part of the analyses performed because of the consensus that not all effects may be valued in monetary terms, thus making cost-benefit analysis difficult if not impossible. Things like pain, disablement and poor health are not easily captured in monetary terms. Most economists would agree that some things are better left unmonetarized, like life as such. We support the opinion that health effects in the denominator of the C/E ratio should exclusively incorporate changes in health status from the perspective of the patient. Thus CEA implies the comparison of all relevant costs with the unmonetarized direct health related effects on quality of life they 'buy'. Loss of income or other non-health related events may affect

general wellbeing, but the objective of *health related quality of life* measurement in economic evaluation of health care is not to capture such effects on wellbeing. Therefore, the opinion of the Panel, that productivity costs should be included in the denominator, along with intrinsic health effects, is not shared by us. The effect-side of the C/E-ratio should be for intrinsic health effects only and income is not believed to be part of health related quality of life.

Also, practice of CEA clearly shows that there seems to be consensus not to include income in quality of life measurement. Generic quality of life questionnaires like SF-36, NHP and EuroQol/EQ-5D avoid questions which allude to effects on wellbeing related to changes in income. The Panel does have a point when arguing that income losses *may* creep in when quality of life valuations are assigned to (different levels of) disability, since such income effects are not explicitly excluded. However, to what extent this effect is measured remains unclear. The implicit exclusion of the influence of income changes may suggest that this influence is small or may even not exist. However, explicit exclusion of income-related effects would be preferable to assure measuring only health related quality of life.

The next two sections are aimed at demonstrating that even if one would want to measure productivity costs through quality of life measurement this will lead to inconsistencies.

Productivity, income and quality of life change from the patient's viewpoint
First, it is clear that a person estimating a change in income will compare income before productivity decrease with income after productivity decrease. What is measured in a quality of life measurement in relation to a decrease in productivity is presented in figure 4.2.

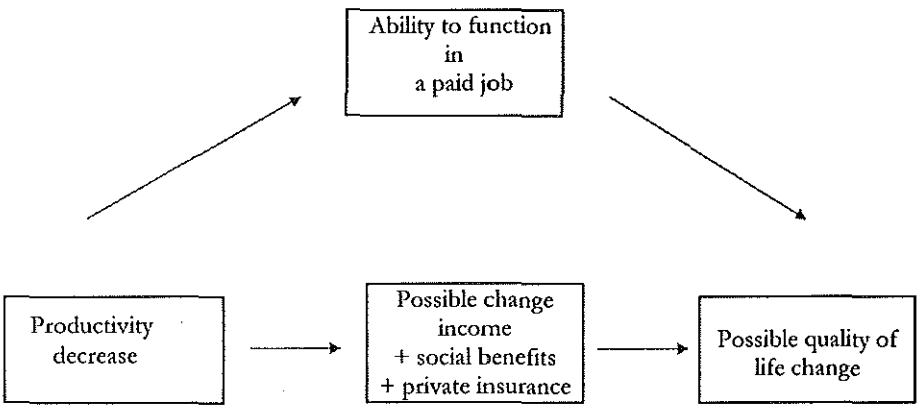

Figure 4.2: Relationship between productivity and quality of life patient.

It is rather obvious that in a quality of life measurement only the effect of total income changes may creep in, while those total income changes may not be related with productivity changes at all. This holds because the increase in social benefits received may partially or totally compensate for the reduction in labor income. This points to a very important role for the level of social benefits in a country when estimating productivity costs this way. If income influences the process of valuation of disability in terms of quality of life, its influence will be highly dependent on social security systems.

From figure 4.2 it may be clear that the results of measurement may differ between countries and even between persons, because of differences in the level of social benefits between those countries or the level of private insurance between persons. International and interpersonal comparisons of cost-effectiveness analyses become much more difficult if we allow social benefits and private insurance to influence the results. It would imply, that all results would have to be reconsidered in the light of social security benefits and private insurance benefits received by the respondents in the sample and that the influence of those aspects would have to be made explicit.

The relationship as presented in figure 4.2 makes the possibility to measure productivity costs through quality of life in a reliable way rather questionable. If we would consider reduced productivity at work it becomes even more obvious that the relationship between productivity loss and quality of life is not that clear. In this case consequent productivity loss would generally not induce an income change and will therefore not be picked up in quality of life measurement at all.

Overall it may be concluded that productivity costs seen from the patient's perspective cannot be measured through quality of life measurement in a reliable way. The latter should only capture the health related effect of not being able to be as productive as before the incidence of illness.

Objections from a societal point of view

First of all, if we agree that a reduction in income may lead to a reduction in quality of life, then it is likely that an increase in income will lead to an increase in quality of life[14]. The Panel clearly states that the societal perspective 'is the only perspective that never counts as a gain what is really someone else's loss' (Gold et al., 1996, page 7). One might also say that we should not count as a loss what is really someone else's gain. In case of replacement of sick workers, if one person endures an income reduction and another person gains a similar amount as his income increases, it will be safe to say that part (or perhaps all) of the quality of life reduction of the first person will be compensated for by a quality of life increase of the second person. From a societal point of view this counter-effect (of someone gaining income and therefore quality of life) should also be taken into consideration. It is much like a zero-sum game, although we do admit that the increase in quality of life does not necessarily have to equal the decrease in quality of life due to income loss. To make the U.S. Panel approach methodologically correct, it would be necessary to measure the gain in quality of life of the person replacing the absentee. Effects of income changes on quality of life should therefore be excluded from health related quality of life measurement. Only the lost ability to work, the true intrinsic health-component, cannot be transferred from the disabled to the person replacing him, since the latter was already able to work. This ability to work or to function in society, or in short to live life like one would want to, should be focused on in health related quality of life.

[14] We would prefer the term 'wellbeing' in stead of quality of life in relation to income, however, we follow the Panel's terminology on this subject in order to avoid confusion.

Secondly, focusing on income from the patient's point of view rather than considering the gains and losses to society leads to problems when considering the time-span over which the reduction in income is incorporated in cost-effectiveness analysis. From a societal perspective only production losses that occur within the friction period should be considered. However, patients will consider the time-span over which income losses are incurred relevant to *them* as opposed to what is relevant to *society*, thus making the outcomes of productivity costs measurement through quality of life even more questionable.

4.5 Conclusions and recommendations

From a societal point of view it is important to analyze the actual production losses due to illness, disability and death and the possible replacement costs. In order to value those production losses it is necessary to calculate the lost value added due to reduced productivity. If after a period of production losses, production again increases to the original level because of recovery or replacement of the sick person, direct productivity costs are no longer incurred. The longer term economic consequences of higher social security premiums, changed wage-productivity ratio's et cetera may be accounted for in a macro-econometric model. The production costs can therefore best be calculated with the friction cost method (Koopmanschap and Van Ineveld, 1992).

From a societal perspective it is necessary to avoid taking into account individual losses compensated by similar gains for other individuals. However, for impacts on health and the ability for role-functioning this is not the case.

Measurement of productivity costs through quality of life will not lead to valid estimations of the real societal costs involved. Therefore in our opinion the recommendations of the U.S. Panel on this subject should not be followed.

Finally, alternative recommendations are presented for including productivity costs in cost-effectiveness analysis, that follow the arguments provided in this chapter. First, productivity costs should be measured from a societal viewpoint rather than an individual one. Only in case of pure health effects, the individual and societal viewpoint lead to the same result. Second, reduced

quality of life because of not being able to engage in work or leisure activities should be counted as health effects rather than as productivity costs. The term productivity costs suggests that these are costs, expressed in monetary terms and not health-effects, expressed in nonmonetary terms. Third, costs of lost productivity have to be calculated with methods taking into account the scarcity of labor and the possibility of replacement, like the friction cost method. Productivity costs should reflect the full costs of lost production from a societal perspective, which cannot be captured in the quality of life measure in the denominator of a C/E ratio as suggested by the Panel.

5. Debate on the valuation of productivity costs

Summary

This chapter presents some of the debate following our response to the recommendations of the US Panel on Cost-Effectiveness in Health and Medicine. In this chapter we primarily concentrate on what Weinstein et al. (1997) call the 'major disagreement' between the Erasmus group and the U.S. Panel, which concerns the measurement of productivity losses during illness. In this chapter the consequences for the individual, for the employer and for the rest of society are considered and it is argued that when following the Panel's propositions for measuring these consequences major theoretical and practical difficulties are encountered. Also, this chapter will further discuss the criticism of Liljas (1998) on the friction cost method which is advocated as an alternative for human capital and QALY-estimations of productivity costs.

5.1 Introduction

In chapters 3 and 4 our recommendations for estimating productivity and patient time costs within economic evaluations of health care were discussed. These recommendations have been criticized, especially the recommendation on using the friction cost method for estimating productivity costs. Weinstein et al. (1997) responded on behalf of the US Panel for Cost-Effectiveness in Health and Medicine to chapter 4, claiming that following our approach would lead to omissions and double-counting of relevant costs. We would omit costs related to changes in leisure time by individuals who are ill and by individuals who are recruited out of the unemployed labor pool. The individual whose health is affected would not be included as a full member of society in our approach and the effects on consumption of others in society would be neglected.

Liljas (1998) also criticizes the friction cost method, but not entirely for the same reason as Weinstein et al. (1997). Liljas argues that the human capital method is the best available method for estimating productivity losses due to illness from a societal perspective. However, as explained below, using the human capital method is not recommendable, especially for longer absence spells. The human capital method serves well to represent ideal individualistic behaviour, but is not intended nor useful to estimate the real value of production losses *from a societal viewpoint.*

This chapter presents adapted versions of our replies to Weinstein et al. and Liljas. In section 5.2 the response to Weinstein et al. is presented focusing on the consequences for the individual patient of productivity losses as well as the consequences for other consumers and the employer. Section 5.3 presents our response to Liljas in which we further explain the friction cost method and give some points of concern on using the human capital method in economic evaluations of health care. Section 5.4 concludes this chapter.

5.2 Response to Weinstein and colleagues

In their response, Weinstein et al. (1997) point out that there is one major difference between the Panel's and our approach which concerns productivity losses during an illness. The effects of productivity losses are segregated into three parts: consequences for the individual patient, consequences for other consumers and consequences for the employer. In this response, we follow this segregation to further explain our recommendations on the subject of productivity costs.

Productivity costs

When discussing the measurement of productivity costs, it is important to understand that production and income are related quantities, both in theory reflecting the value added of productive persons, or their contribution to overall wealth. Our approach is to measure this quantity from the production side, calculating lost production and/or increased production costs and medium term macro-economic consequences due to disease (Koopmanschap, 1995), while the Panel takes income measured through quality of life as a proxy for this quantity. However, as demonstrated in our original response, individual income may only have a weak link with production change, particularly due to social or private insurance. By taking the production side as

starting point, we are by no means failing to incorporate the patient as a full member of society, but we attempt to balance the production loss due to the absence of the patient with the production 'gain' of the previously unemployed, replacing the former after a friction period and other changes in society (including costs to the employer).

Patient as a member of society.
If the lost value added as a consequence of disease is measured from the production side, the only remaining impact on the patient is the change in wellbeing of the patient due to his or her lost ability to work, which is generally measured in health status questionnaires. The problems with incorporating income changes in quality of life measurement have been discussed extensively in our original response (Brouwer et al., 1997a) and recently Johannesson also expressed some doubts on incorporating income in quality of life (Johannesson, 1997). These arguments are not discussed by Weinstein et al. (1997) but strongly suggest that when income is used as a proxy for added value and measured in terms of quality of life, there are numerous theoretical and practical problems to overcome. Most notably, income may only have a weak link with productivity, which makes the incorporation of income in QALY measurement as proxy for productivity losses rather dubious. Some productivity changes may occur completely without any change of individual income, as social insurances and sickness benefits often will keep individual income on the original level. Also, after a *friction period* society may have adapted to the absence of a person by redistributing the original workload over others in society, e.g. formerly unemployed. After this, society will no longer incur production losses (Koopmanschap, 1995), however, at that moment in time, the affected patient may still experience a lower income. Taking this into account for as long as the patient experiences this, without also incorporating counter-changes in society (e.g. income increase and thus QALY-increase of a formerly unemployed) may lead to a serious misrepresentation of *societal* production losses.

The employer
Secondly, Weinstein et al. (1997) mention the impact of longer absence of an employee on the employer. As indicated, the transient costs of replacing a worker should not be forgotten. Costs related to advertising, training, etcetera are indeed all real societal costs which fall on the employer and we agree that these additional replacement costs should be incorporated. To define only

these costs as friction costs makes friction costs a very limited concept. The friction cost *method* as proposed does not merely capture these additional costs to the employer but also the lost production during the friction period itself as a consequence of absence or disability. We feel that measuring production loss in an integrated way together with these additional costs of selection and training is more reliable than using 'transient costs' and income to the individual as constituting total production loss.

Others in society

Weinstein et al. (1997) argue that costs for other consumers in society should not be neglected, although they do not offer a method for measuring these costs. We calculated the medium term macro-economic consequences of productivity changes in society induced by disease in terms of changes in national income (Koopmanschap, 1995). These concern changes in the redistributive system, i.e. higher premiums for sickness benefits and disability benefits, and a lower labor supply. These changes will subsequently influence labor costs per unit of output and therefore the competitiveness in the sector exposed to international competition and through this again the economy's production. These multiplier effects can only be calculated using an econometric model of a nation's economy and for the Netherlands they were estimated to be 0.8% of GDP in case of total eradication of illness (Koopmanschap, 1995). Since short-term friction costs account for the major part of total costs and since medium term macro-economic consequences are difficult to calculate and may be subject to changes in social security system, one may neglect these costs when considering single health care interventions with a relatively minor effect on total productivity in a country. In a sensitivity analysis the possible importance of medium-term productivity costs may be further investigated.

Leisure time

Another point Weinstein et al. (1997) raise is leisure time. Although we did not discuss changes in leisure time in our original paper, where we focused on productivity costs related to paid labor, we agree that leisure time indeed should not be forgotten in a cost-effectiveness analysis. In our opinion in cost-effectiveness analysis the use of QALYs provides us with an ideal instrument to incorporate leisure time. When respondents assess quality of life, it is recommendable to also let them consider the influence of reduced possibilities to engage in leisure-activities. Johannesson recently expressed a similar opinion (Johannesson, 1997). Weinstein et al. argue that we would

leave out the costs of reduced leisure of the previously unemployed who replaces the ill individual. If a patient becomes unemployed because of illness and another person takes his job after the friction period, the absolute amount of leisure time remains the same, since the patient now has more leisure time and the newly employed has less. The obvious difference in the capability to fully enjoy this leisure time is captured in the QALY-measure as a QoL-loss. Secondly, the definition of unemployment in many countries involves *involuntary unemployment*. This indicates that unemployed may be more than willing to sacrifice some of their leisure time in order to work for a pay. The valuation of the sacrificed leisure time therefore may not be that high. Micro-economic approaches showing that leisure is only engaged until the point where its marginal utility is equal to the wage rate obviously do not fully apply when the person involved is externally restricted in his choice. The desire to work of many unemployed seems to indicate that the difference in valuation of leisure and working time is rather significant.

Conclusion

In chapter 4 (Brouwer et al., 1997a) we tried to demonstrate that if one would want to incorporate income in QALYs as a proxy for productivity costs, the results from such measurement would most probably misrepresent the quantity one would want to measure, which is the lost value added to society. So, in contrast to the assertion of Weinstein et al. (1997), we do not accuse the Panel of ignoring 'indirect costs' or productivity costs (we support this new term). Our main concerns with measuring productivity costs through quality of life are the following:

1. Income should not enter quality of life measurement and current practice shows a consensus to avoid questions alluding to income. Real costs should be incorporated exclusively in the numerator of the C/E-ratio and pure health effects should exclusively be incorporated in the denominator of the C/E-ratio.

2. Individual income may only have a weak link with production change, which can be measured more reliably at the production side.

3. Point 2 is even more noticeable in case of reduced productivity at work. This may not induce any income change at all, while it does constitute a real societal loss. These losses should not go unmeasured as they would following the Panel's approach.

4. The income increase of the previously unemployed replacing the sick individual is not considered by the Panel. Therefore, the Panel abandons the societal perspective by not incorporating this individual as a full member of society. Yet, if income decrease will induce a QALY-loss, then an income increase will obviously induce a QALY-gain.

5. Related to the previous point, one can say that patients will consider income losses (if asked to do so) as long as these affect them, while on a societal level, total income (or rather production) may be restored after replacement. The last two points indicate why we categorized the Panel's approach as taking an individual rather than the societal viewpoint.

5.3. A reply to Liljas (1998)

In this section Liljas' criticism of the friction cost method is considered and his alternative, i.e. using the traditional human capital methos, is critically discussed.

Critique on friction-cost method
Liljas (1998) discusses the friction cost method at length in an attempt to demonstrate flaws in this method. His main argument is that the friction cost method is not embedded in neo-classical welfare economics. We will discuss some of Liljas' specific arguments below.

Short-term absence
Liljas points out that using the friction cost method would violate one of the 'fundamental' axioms of neo-classical theory, i.e. that firms will employ labor up to the point that marginal costs equal marginal benefits. Several responses are possible to this criticism. First, reality does not coincide with neo-classical models of behavior. There is no perfect competition between firms in the marketplace and all kinds of managerial objectives may be incorporated in the optimalisation behavior of the firm that will yield results not in line with standard textbook situations. Secondly, the friction cost method does not correct the *wage rate* for the elasticity of time versus output, but it corrects the average value added per worker (as it focuses on productivity), reflecting the value of commodities and services produced (Koopmanschap et al., 1995). It

is not unreasonable to assume that 80% of average value added would be close to marginal labor costs.

Liljas does not separate some of our remarks about internal labor supply in relation to high transaction costs from our final calculations. In the base case calculations of Koopmanschap et al. (1995) all costs of short-term absence from work were counted. Hence, in the base case we *did not correct* for internal labor supplies, nor for making up for lost work after illness nor for canceling less important work, because we did not dispose of reliable estimates of the magnitude of these phenomena. However, these phenomena may prove to be substantial (e.g., Severens et al., 1998), although more research is needed.

Longer term absence

For longer absence spells Liljas seems to imply that the permanent absence of one person will induce a series of friction periods. However, Liljas seems to forget that when somebody is already employed, he normally has to give his employer notice through which the latter may arrange replacement in the meanwhile. If replacement is then still not possible according to Liljas, then it is strange why even the human capital method stops counting at retirement, since even retirement would induce an infinite series of replacement friction periods in such reasoning. Furthermore, it is not claimed that all workers are replaced by formerly unemployed, and friction cost estimations do rise with a lower unemployment. It may be true that in case of a chain of replacements the production loss and training costs may be somewhat larger than in case of one replacement, but it is probably not proportional to the number of changes in the chain. Also, when unemployment is very low perhaps all absence should be considered to lead to productivity costs, but not permanent disability and premature death (Koopmanschap and Rutten, 1996b).

Liljas presents the human capital method as the alternative for the friction cost method. This method is embedded in neo-classical economics, which should imply theoretical superiority, according to Liljas. It should be noted, that the micro-economic oriented human capital method serves well to represent *ideal individualistic behaviour*, but is not intended to estimate production losses from a *societal viewpoint*. In our opinion it is undeniable that replacement mechanisms do in fact occur and that there is involuntary unemployment.

For long term absence from work, disability and mortality the human capital method still overestimates the true productivity costs of illness, by denying any adaptation mechanism for organisations as a response to absence of workers. Consequently, this method only estimates potential, but not realistic productivity costs, continuing for decades, denying the dynamics in the real economy of labor substituting labor and capital substituting labor. Is it realistic to claim that if a paid employee dies at age 30 in an accident, society will suffer production losses up till his retirement age of 65, i.e. 35 years? And if someone aged 45 becomes permanently disabled, is it realistic to claim that society will suffer production losses for 20 years? In fact, with this reasoning Liljas implies that there is permanent full employment all over the world and absolutely no possibility of replacing sick workers. Granted, unemployment may be low at some points in time, but it is highly implausible to assume a state of permanent full employment.

Liljas states that the friction cost method implies 'that it would be possible to solve the problems of unemployment in society by reducing the number of hours worked by employed and having unemployed workers work these hours instead ... something that seems very unrealistic'. (Liljas, 1998) First of all, this is a clearly misuse of the friction cost method which is not intended at solving unemployment but at estimating indirect costs. Secondly, there is an obvious difference between assessing the indirect costs (and benefits) of a certain marginal health care program and assessing ways to totally eradicate unemployment. Of course, there are many institutional problems that interfere with such a mechanism to fully work, but for marginal changes indeed such replacement will take place. Unemployed workers are not unable to replace 'normal' workers. Indeed, Liljas seems to imply that the market selects the best workers and leaves 'second grade workers' unemployed and therefore, a reduction in the working hours of employed cannot be made up for by hiring unemployed persons. We would not generally classify the unemployed as being inferior in a productive or any other sense. Indeed there are institutional barriers, educational barriers, and perhaps even social barriers that prevent the labor market to function perfectly flexible, but it is denied that (at the margin) unemployed persons cannot replace employees who fall ill.

Leisure time

Liljas' correctly remarks that leisure time should not be neglected in economic evaluation of health care. However, this issue has already been discussed on

several occasions and it needs to be stressed that it seems the most appropriate to incorporate the reduced ability to enjoy one's leisure time to the fullest in an economic evaluation in terms of QALYs (see chapters 3 and 4). This point of view is also advocated by Johannesson (1997). Although this solution is not without problems it is considered to be the best available. Also, it means that the friction cost method does not have to deal with leisure time and *a fortiori* should not deal with it to avoid double-counting.

Direct labor costs

Liljas reaches the conclusion that the opportunity costs of replacing an ill employee are close to zero in the friction cost method. This would be the case, because no costs are attached to the fact that a formerly unemployed substitutes leisure time for time spent on working for pay. Liljas goes on by stating that if the opportunity costs of working are indeed close to zero, as he suggests is the case in the friction cost method, then one should value labor needed for a health care intervention quite 'close to zero' as well. Since the latter is obviously unacceptable for most economists, this should demonstrate the inappropriateness of the friction costs method.

It is noted here that the friction cost method is not a full economic evaluation method in itself, but merely part of a complete analysis. Another part of the analysis is QALYs and it is stressed that it is in terms of QALYs that leisure time is captured. When a formerly unemployed gives up labor to replace an ill employee, it should be noted that in principle the full amount of leisure time remains the same on a societal level, only it is the previously working patient that now has more leisure and the formerly unemployed less. The difference in the possibility to enjoy leisure is assumed to be fully captured in the QALY-measure.

The equaling out of leisure time is a specific feature when dealing with replacement (due to illness) and therefore, it should not be used for additional work as done by Liljas.

Conclusions

Liljas has tried to advocate the use of the human capital method for estimating indirect costs instead of the friction cost method. It needs to be stressed that neither method is perfect or captures all indirect costs in a straightforward manner. The human capital method serves well to represent ideal individualistic behaviour, but not to estimate societal production losses.

The claim that the human capital method is better because it is embedded in neo-classical theory may be easily countered by stating that the real economy clearly does not resemble the neo-classical model. Since we probably do not disagree on that, we should find ways to come up with methods that estimate indirect costs in a more realistic way. The friction cost method, though not perfect, intends to do just that.

5.4 Concluding remarks

The area of productivity costs measurement and time costs remains a controversial one. However, hopefully, discussions like the ones presented in this chapter contribute to a better understanding of each other's viewpoint and ultimately to a more uniform methodology on the estimation of productivity costs in economic evaluations of health care.

6. Productivity losses without absence: measurement validation and empirical evidence

Summary

Productivity losses without absence are scarcely discussed in the literature. In this chapter the construct validity of three different measurement instruments for productivity losses without absence is investigated. The data was collected under employees of a Dutch trade-firm, not in specific patients groups. On an average day, over 7% of the respondents were working with health problems, indicating that productivity losses without absence is a quite common problem. The amount of production losses related to these health problems are relatively small. However, for specific patient groups the costs related to these productivity losses may be substantial.

6.1 Introduction

In economic evaluations of health care interventions indirect non-medical costs or productivity costs often play an important role when the prevalence of the illness involved concerns people with paid or unpaid work. The incorporation of these costs in economic evaluations has been much debated. Both questions whether they should be counted (e.g. Russell, 1986; Gerard and Mooney, 1993) and if so how they should be estimated (e.g. Koopmanschap et al., 1995; Gold et al., 1996; Brouwer et al., 1997a) have been extensively discussed in the literature. Recently the consensus for incorporating these societal costs seems to be growing (e.g. Gold et al., 1996; Drummond et al., 1997). Attention is usually focused on costs related to absence from paid work. However, it is obvious that absence from paid work is not the only situation causing production losses related to disease. Besides impaired ability to perform unpaid work people may be at work while not being in optimal health. Mild (chronic) diseases or the onset of acute infectious diseases are examples of situations in which people may not function to their normal ability, yet are not impaired 'enough' to stay at home.

These productivity losses without absence are potentially important, yet almost neglected in economic evaluations of health care.

Only two attempts to estimate the costs related to this productivity drop were found in the literature. One was developed by Osterhaus et al. (1992) and later adapted by Van Roijen et al. (1996) and the other was developed by Van Roijen et al. (1995a). The adapted Osterhaus method (O-method) and that of Van Roijen (VR-method) were compared in estimating productivity costs without absence due to migraine and showed substantially different estimates, i.e. 968 versus 277 million Dutch guilders respectively (1995). This shows these costs can be substantial for certain diseases and that more investigation is needed on how to realistically estimate these costs.

In this chapter, we compare the O-method, the VR-method and an alternative experimental method, which aims at measuring the Quantity *and Quality* of work performed on a daily basis, the QQ-method. We investigate the construct validity of these methods using data of employees of a Dutch trade firm. Section 6.2 presents a brief introduction of the three methods. In section 6.3 the instruments and data used in this study are discussed. Section 6.4 discusses the results of the three methods, while section 6.5 concludes this chapter.

6.2 Brief introduction of the three measurement methods[15]

The adapted O-method for determining the efficiency losses due to illness, as used in the Dutch Health and Labor-questionnaire, consists of two questions. We have chosen to use this 'adjusted O-method' to be able to compare the outcomes with the earlier comparisons (Van Roijen et al., 1995a; Van Roijen et al., 1996). The O-method asks how many days during the *last two weeks* one went to work while suffering from health problems and on a VAS-scale the average *efficiency* on these days. A difficulty in this method is that the efficiency indication may not necessarily be interpreted as the amount of work respondents did compared with normal, but rather what they *could do* compared with normal, although it is not clear how subtle this difference is. If full utilization of labor time is not required on average (i.e. one does not work at full capacity at all times), using the average efficiency-score may overestimate production losses. Overestimating production losses may

[15] The three measurement methods as used in the study are fully printed in appendix A.

additionally occur since the possibility of making up for lost work during regular hours without creating new production losses is ignored. The term efficiency may induce respondents to consider quality losses as well, but it is unsure to what extent these contribute to the final score.

The VR-method also asks people to indicate on how many days they were less able to perform their work over the *last two weeks* and how many extra hours they would have had to work to make up for the work they lost during the days feeling ill at work. This should provide an indication about how much production was actually lost. The indicated number of hours needed to make up for lost production enables a monetary valuation ('replacement-value'). Here productivity losses that can be recovered during normal working hours do not count as production losses, since only *extra* hours are asked for here. Shifting production from one day (with health problems) to another day (without or with less health problems) is thus seen as costless recovering of lost work. The VR-method may underestimate production losses, however, if not all work can be made up for later on. E.g., redistribution of work over colleagues ensures that the collective workload is finished on time because colleagues make up for lost work during regular hours and some types of work cannot be made up for, for instance in a production line. Still, these two situations do not have to be costless. Redistribution of work reveals some slack in the organization or working extra time of some employees (not necessarily the person less productive) which is not without costs and neither are unrecoverable production losses. These situations may cause missing values or zero-answers in the VR-method, while they do indicate impaired productivity on the O-method. An additional problem in interpreting VR-answers is that it is unclear if these extra hours indicated represent hours worked at normal performance level or the reduced level that respondents may be still experiencing. Consequently, the answer is a conservative estimation of production losses, but still difficult to interpret exactly.

The QQ-method was developed in an attempt to more precisely measure the consequences of illness while working. First, on a *daily basis* respondents are asked to indicate on a VAS-scale from 1 to 10 how much work they actually performed during regular hours compared with normal (Quantity scale). It is a question about output rather than efficiency, indicating how much work had to be made up for *during overtime or regular work time* or was lost. Translating the answer into costs again has the possibility of overestimating true costs

involved if it is possible to make up for lost work during regular hours. Fluctuations in daily performance level are explicit in this method.

Second, the quality of the work performed on a day is indicated on a VAS-scale from 1 to 10 (Quality-scale). One can imagine that not all losses are quantitatively expressible and a quality indication therefore may provide valuable additional information. However, the quality scale raises new questions as well, especially on how to treat answers on this scale. Straightforward multiplication of the quantity and quality components may provide an indication of total performance, translating qualitative into quantitative losses. However, although this combination seems adequate as a first attempt to combine qualitative and quantitative losses, it remains unclear whether it is the correct way. Still, it should be noted that, for some types of work, qualitative production losses seem more relevant than quantitative losses, e.g. for school teaching.

Comparing the three methods described above, it seems that the O-method measures the production capacity of the respondent, leading to a maximum indication of production losses. The VR-method measures the amount of recoverable lost production, that has not yet been made up for. This appears to provide a minimum estimate of production losses. The quantity-scale of the QQ-method is expected to yield estimates in between that of the O- and VR-method.

6.3 Methods

We aimed at establishing construct validity of the three measurement methods, basing our hypothesis on the expectations derived from section 6.2 and previous experience. Construct validation is tested by examining whether different methods measure the underlying quantity (production losses) in expected ways. Here, it is expected that the VR-method will yield the lowest results in terms of hours lost (with a relatively high number of 'zero'-answers), the O-method would yield the highest results and the results from the quantity scale of the QQ method would lie in between these two estimates, more close perhaps to O-results than to VR-results. Furthermore, the quality scale of the QQ-method was used to estimate qualitative losses due to illness at work.

The data were collected in a Dutch trade firm. The questionnaire 'Consequences of Illness' was sent to all employees of this firm. In this diary people were asked to indicate for each day of one week whether they had been less productive at work due to health complaints. Respondents could indicate on two different scales how the quantity of work performed related to normal and how the quality of the work performed related to normal (QQ-method). At the end of the week people were asked to answer the questions of the O-method and the VR-method (see Appendix A). The latter questions allude to the period of one week only, to make the answers comparable to the daily answers. A gold standard, i.e. individual data on daily production, to compare the different answers with, was unavailable due to anonymity of respondents. This is why validation on the basis of construct validity was chosen.

The total number of employees was 216. The relevant questions were sent to all employees 5 times. The diaries were sent 4 times in the months June and July 1997 with one week in between each weekly diary. The fifth diary was sent approximately three weeks later in August 1997. The total number of returned diaries was 543, implying an overall response rate of 50%. Due to the anonymity of respondents we could not further investigate the non-response.

One of the primary aims was to establish how many people indicated that they were less productive at work on an average day. This should give some indication of the importance of the phenomenon productivity losses without absence. Since all employees were potential respondents, all types of work were included in the study, however, most of the respondents were engaged in work that was automatically replaced by colleagues and could not be made up for the next day. In Table 6.1 some characteristics of the respondents of the diary are presented.

Consequences of Illness:
Male respondents: 117
Female respondents: 416
Blanks: 10
Average age males: 35.0 years
Average age females: 40.0 years

Table 6.1: Main characteristics of the respondents.

6.4 Results from the diary 'Consequences of Illness'

Of the 543 diaries 'Consequences of Illness' returned, there were 57 in which it was indicated that at least one day during the week the respondent was feeling ill while being at work. This means 10.5% of the diaries were (in principle) useful for this study. After investigation it was concluded that 53 diaries were consistently completed and could therefore be used in the analysis below. Analyzing this subgroup of 53 diaries, it turned out that the average age (42.2 years) was significantly higher than that of the other respondents (those not ill), which was 38.5 years ($p=0.0008$) and that the group had a significantly higher proportion of women ($p=0.0278$); 87% versus 77% .

Analyzing all 543 questionnaires it could be concluded that on an average working day 7.35% of the respondents indicated that they felt ill while being at work (minimum over 5 measurement moments 5.43%; maximum 9.49%). Just to give an indication of the commonness of the problem (at least in this firm), in the same sample the self-reported absence due to illness was 2.85%. This absence rate is quite low and lower than the average absence rate the company registered itself, which was on average 4.5% for the relevant months (never below 4.15%), indicating that a relatively high amount of non-response may be from people that were absent from work more frequently. As stated above, we were unable to further investigate the non-response due to the anonymity of respondents.

The *impact* on production of productivity losses without absence is obviously smaller than the average of 7.35% of employees feeling ill at work might suggest, since the affected employees loose only a fraction of their normal production, as shown in Table 6.2 (where 'Quantity scale' represents the quantity scale of the QQ-method).

Method	Average scale answer (st. dev)	Average hours lost (st. dev)	Number of 'zero' hours lost	Total number of hours lost	As % of Total hours ill at work	As % of total hours firm level
Osterhaus	8.77 (1.14)	1.72 (1.92)	15	91.4	6.32%	0.59%
Van Roijen	-	0.23 (0.73)	45	12.3	0.85%	0.08%
Quantity scale	8.93 (1.16)	1.54 (2.00)	20	81.6	5.64%	0.52%

Table 6.2: Average scale-scores, numbers of hours lost and relative impact according to the three methods.

In the diary the respondents could tick one of the 6 specified disease-categories or tick a box for the category 'other complaints'. Table 6.3 presents the results for the six disease categories that were at least indicated on one day.

Causes of illness	**Frequency in days**
Cold or Influenza	25
Back or Neckproblems	8
Migraine or headache	17
Psychological complaints /stress	15
Stomach/Intestines	7
Other complaints or combinations of complaints	42 (25 combinations)
Of which:	
Allergy	8
Flue/cold and Headache/migraine	5
Headache/migraine and Psych. Complaints/stress	7

Table 6.3: Causes of illness at work and frequency in days.

Comparing the three methods: construct validity

If one examines what is being asked in the VR-method and the O-method, it is clear that the answers of the respondents cannot be compared in a straightforward way. This is because in the Osterhaus method the number of working hours is also important. A score of 9 on the efficiency scale will be interpreted here as a loss of 0.1 * the normal hours of work per day. So, if a

person indicates that he has been less productive for three days in a week and normally he works 8 hours per day and his efficiency score was 0.9 the total loss will be calculated as: (1 - 9/10) * 8 hours * 3 days = 2.4 hours lost. When comparing the answers of the O-method and the VR-method including such a correction for the number of hours worked, one can detect a correlation between the two variables, although they are far from equal. Figure 6.1 presents the results, both for the VR-method and the O-method. The respondents are ordered by their score on the O-method.

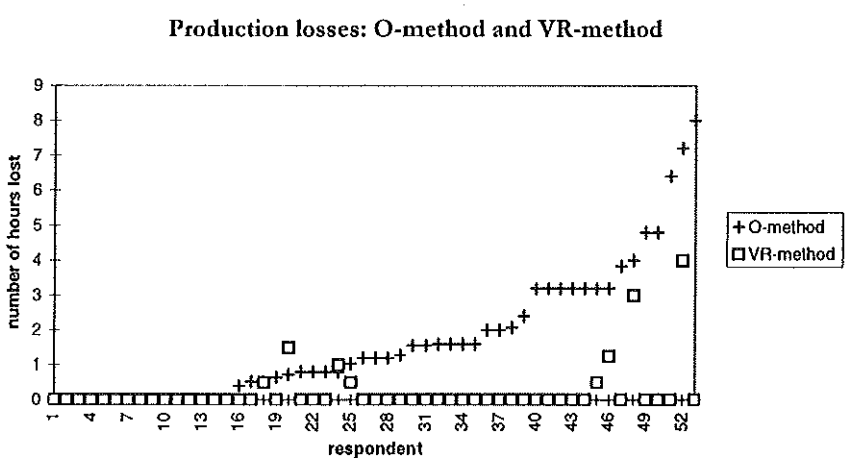

Figure 6.1: Relation between the answers of the O-method and those of the VR-method.

A first observation may be that 15 people indicated zero hours lost in both methods. For the VR-method alone, this number is 45, implying that nearly 85% of the respondents feeling ill at work indicated that they would need zero extra hours to make up for lost work. The (Pearson) correlation coefficient between the results is 0.38, which is close to the 0.41 reported by Van Roijen et al. (1996). In a linear regression analysis the relation between the results from both methods turned out to be significant. The results from that analysis (equation number 1) and all other regression analyses are presented in Table 6.4 (where *b* represents the coefficient, *a* represents the constant and Qt represents the score on the quantity scale of the QQ-method).

Equation Number	Form	Number of Observations	b (t-value)	a	Multiple R	R-square
1	VR = b*O + a	53	0.15 (2.96)	-0.02	0.38	0.15
2	VR = b*O + a	8	0.47 (3.62)	0.33	0.83	0.69
3	O = b*Qt + a	53	0.89 (16.94)	0.36	0.92	0.85
4	O = b*QQ + a	53	0.55 (17.33)	0.23	0.92	0.85

Table 6.4: Results for several regression equations, relating production losses according to four methods.

The results from regression number 1 indicate that, since b is smaller than 1, i.e. 0.15, when the Osterhaus calculation of hours lost increases with 1 hour, the Van Roijen method increases with 0.15 hours (9 minutes). A t-test confirmed that O-results are significantly higher than VR-results (p-value<0.0001). This evidence supports the prediction made in the previous section. Because including zero-answers to the VR-method in this analysis might be considered comparing two different regimes (since zero does not have to mean no production losses as explained in section 6.2), we also performed the analysis after deleting all observations scoring zero in the VR-method, which ensures only comparing observations in which an estimation of production losses was indeed possible. The resulting regression analysis indicates that the relationship between the results of the two methods becomes stronger when the zero answers on the VR-question are deleted (regression number 2, Table 6.4). This provides some evidence for our theoretical prediction, however, the low number of observations (n=8) inhibits firm conclusions. This result does indicate that the type of work examined may influence the correlation between the two methods. In this study much of the work not performed by someone less productive was automatically replaced by a colleague or unrecoverably lost. The value of *b* however remains smaller than 1 (0.47), implying that marginally VR-results are smaller than O-results.

To further assess the construct validity the results of the VR-method and the O-method have to be compared with those of the Quantity-scale of the QQ-method, the Qt-method. In figure 6.2 all three results are shown.

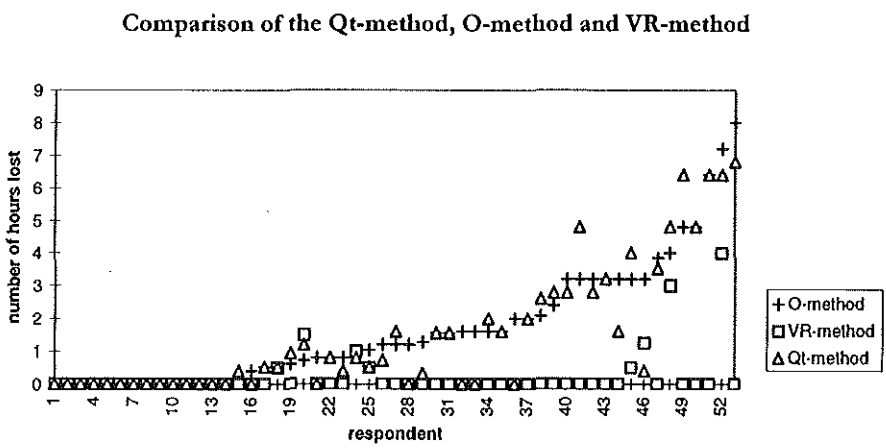

Figure 6.2: Comparing the Qt-method to the O-method and the VR-method.

On average the total amount of hours lost for the Qt-method, O method and VR-method respectively are: 81.6, 91.4 and 12.3, as shown in Table 6.2. On average therefore, the Qt-method yields lower estimates than the O-method, but figure 6.2 demonstrates that in individual cases this may not be the case. If we examine the results from the Qt and the O-method, it turns out that from the 53 observations, 25 times both methods yielded the same result (14 times both methods estimated the number of hours lost as zero). In 17 instances the O-method estimated the production loss higher than the Qt-method, 11 times both methods yielded the same result (not zero) and 11 times Qt-results were higher than O-results, which is not in line with our expectations. The Pearson correlation coefficient was 0.92. The regression analysis, as shown in Table 6.4, number 3, shows that the one hour extra in Qt will lead to 'only' 0.89 hours increase in the O-results. However, the latter will start higher, due to the positive intercept value. To analyze whether the Qt-method estimates production losses significantly lower than the O-method, we performed a one tailed paired two sample t-test on the results. For all 53 observations, the p-value was 0.04594. We performed the same t-test for a sub-group of

respondents, which included the 39 respondents that indeed scored some actual production loss on either scale. By eliminating the zero-zero combinations, it was investigated whether these combinations distort the comparison of the two methods, which are obviously aimed at estimating real production losses. The difference remained just statistically significant ($p = 0.04596$). Two one-tailed two paired sample t-test of the 53 pairs of VR-results and Qt-results and the 40 remaining pairs when deleting all 'zero-zero' combinations confirmed that the Qt-method yields significantly higher results than the VR-method ($p<0.0001$ for both tests). The Pearson correlation coefficient between the two methods was only 0.40, close to the correlation between VR-results and O-results.

One may conclude that, based on the concept of construct validity, there is some empirical evidence to support the hypothesis that Qt would lie in between the O-method and the VR-method for this sample. In relation to the VR-method the results are clear in the sense that the VR-results are obviously lower than Qt-results. The comparison between the Qt and the O-method would benefit from more research using a larger sample.

The QQ-method: incorporating the quality component

As introduced in section 6.2 the complete QQ-method measures both quantity (Qt) and quality (Ql). To give an indication of how these two correlate, the average Qt and Ql answers are shown in figure 6.3. These are the average scores per respondent on both the quality scale as well as the quantity scale.

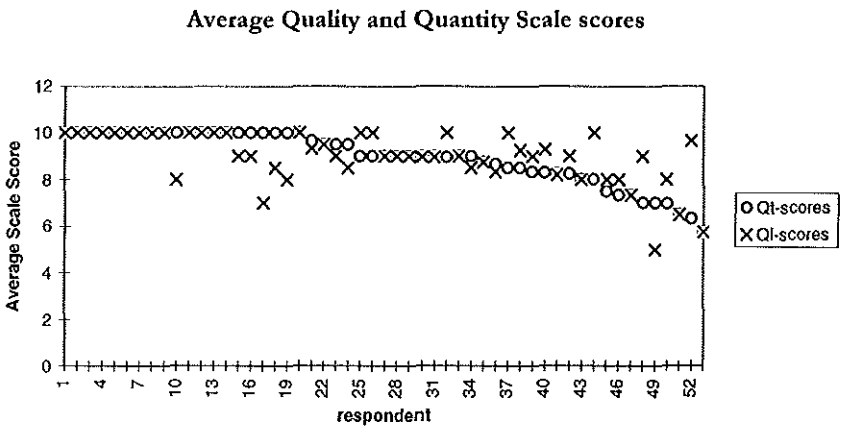

Figure 6.3: Average scores on the Qt-scale and Ql-scale compared

The Pearson correlation coefficient is 0.59 for these two variables. On average the quality score was about equal to the quantity score: 8.97 versus 8.93. As mentioned in section 6.2, it is unclear how these two quantities would result in a total productivity loss. One possible way is a 'straightforward multiplication' of the scale scores, however, there are many other possible combinations. The results from this 'multiplication' are presented in figures 6.4 and 6.5. Figure 6.4 presents the Qt-losses with respect to the Ql-losses, while in figure 6.5 the total QQ-results are compared with the results from the O-method. Table 6.5 contains an exemplary calculation of the final QQ-score. Note that we use the scale-scores linearly to produce the hours of work lost, as was done for the O-method.

Imagine a respondent who has been ill at work for 1 day. His score on the Quantity Scale is 8 and his score on the Quality Scale is 7. This respondents works 8 hours per day.
The Qt-loss is calculated as: (1-8/10) * 8 = 1.60 hours.
This means that we assume that this person effectively worked 8 - 1.6 = 6.4 hours that day. This 6.4 hours of work were performed at 70% of normal quality level. Therefore the Quality loss of the performed work is calculated as: (1-7/10) * 6.4 = 1.92 hours
The final QQ-score now amounts to: 1.60 + 1.92 = 3.52 hours

Table 6.5: Example of the calculation of QQ-scores

In figure 6.4 the quantity and quality losses are compared for the 53 respondents.

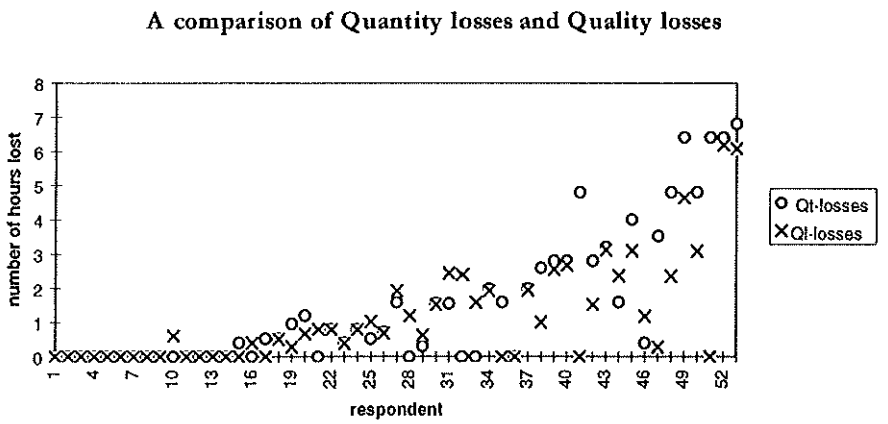

Figure 6.4: A comparison of Qt-results and Ql-results.

The correlation coefficient between Qt and Ql is 0.71. In total Qt losses amount to 81.6 hours and Ql losses to 62.8 hours. In figure 6.5 the total QQ-score, calculated by means of multiplication, is compared with the O-method.

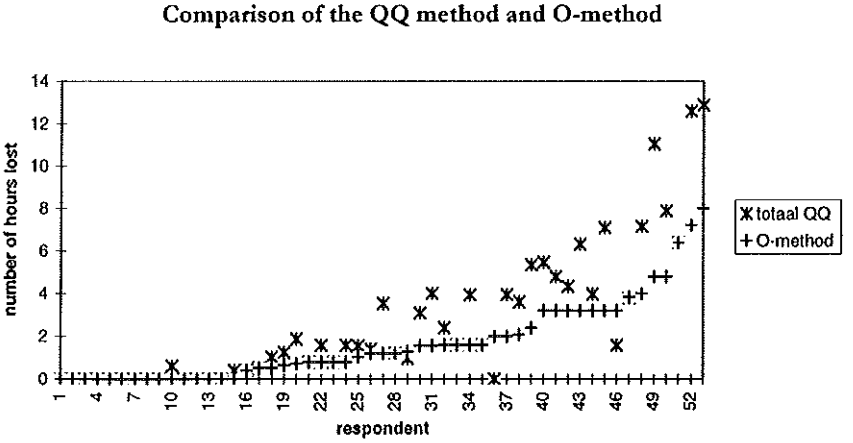

Figure 6.5: Comparing the number of hours lost according to the QQ-method and the O-method.

The correlation between the number of hours lost calculated with the QQ-method and the O-method is 0.92. In a regression analysis, as shown in Table 6.4 equation number 4, the relation turns out significant. Calculating the total impact from illness at work with the QQ-results (i.e. a total of 144.37 hours 'lost'), as already performed in Table 6.2 for the Qt, O and VR-method, the number of hours lost would become 10% of the total amount of hours worked by respondents while being ill. On the total sample or firm level, the impact may be considered modest: 0.93% of total hours worked.

6.5 Discussion

In this chapter we have investigated the construct validity of three measurement methods of productivity losses without absence. Statistical analysis indicated that our hypothesis (Qt-results on average lie in between VR-results and O-results) could not be rejected.

Compared with the Qt-method and the O-method, the VR-method was a consequent outlyer with a relatively high amount of zero-answers. There are two possible explanations for these answers: making up for lost work during regular hours or the impossibility of making up for lost work. The specific type of work of most respondents in this study, in which making up for lost work later is impossible, leads us to suspect that many of the registered zero's in fact may imply that people cannot make up for lost work. If this is indeed the case, VR-estimates of productivity costs without absence may be considered an underestimation, since not being able to make up for work lost (nor internal replacement) may be without cost. This shows the necessity of acquiring additional information in studies like these on the type of work of respondents, type of organization, possibilities for making up for lost work and replacement issues. The exact organisation of work in the firm in this study will certainly have had an impact on (relative) results. Measurement methods should give explicit attention to these aspects.

Both the Qt-method and the O-method ignore the possibility of making up for lost work later during normal hours and therefore may yield an overestimation in terms of costs. The Qt-method provides somewhat more conservative cost-estimations than the O-method and the type of questions

asked may be considered to indicate better the real quantity of work lost. However, more research is needed to compare these methods further, preferably to a gold standard, i.e. daily individual production data. Also, respondents may be asked in more detail about the way they come up with an answer to the questions asked in the different methods.

A noteworthy result from this study is that illness at work is quite common. On an average day over 7% of the respondents indicated to experience health problems while being at work. The total impact from the phenomenon productivity losses without absence at the firm-level, however, is quite low in this study. This is because respondents that were ill at work still could perform most of their work. More research in other firms could provide insight on whether this result is similar to the impact of productivity losses without absence in other firms. Even with the QQ-estimates, this type of production loss is estimated to be maximally 0.93% of total work time in this firm. To give an estimation of the monetary impact, using an estimated value added of 80,000 Dutch guilders, this means a value of lost work time of 774 guilders per year per employee.

Finally, it seems worthwhile to further develop, validate and use these methods in different settings, including the QQ-method, also in specific patient groups. Regarding productivity losses without absence, especially the process of making up for lost work (also by colleagues) and the quality of work should be investigated further. In that sense, the results presented in this chapter are not final answers to how impaired productivity at work should be measured and valued, but give an indication of what elements are of importance, which measurement methods are available at this point and which problems are attached to the different methods.

Acknowledgements

We are grateful to Jaap van Dijk, Marcel Knotter and Leida Lamers for their useful comments. This study was financially supported by Glaxo Wellcome, United Kingdom.

Appendix A

Different methods in 'Consequences of Illness'

The questionnaire 'Consequences of Illness' covers production losses without absence through the QQ-method on a daily basis. On all days the exact Dutch phrasing was:

Op de schaal hieronder kunt u aangeven hoeveel werk u vandaag hebt gedaan in uw normale werktijd ten opzichte van een normale dag. Een 1 omcirkelen betekent dat u niets kon doen en een 10 betekent dat u evenveel als normaal kon doen.

1	2	3	4	5	6	7	8	9	10
Zo goed als niets									**Evenveel als normaal**

Wilt u inschatten hoe de kwaliteit was van het werk dat u vandaag hebt gedaan ten opzichte van normaal. Een 1 omcirkelen betekent dat uw werk van zeer slechte kwaliteit was en een 10 betekent een even hoge kwaliteit als normaal.

1	2	3	4	5	6	7	8	9	10
Erg slechte kwaliteit									**Zelfde kwaliteit als normaal**

Translated into English:

Could you indicate how much work you actually performed today during regular hours as compared with normal on the scale below?

1	2	3	4	5	6	7	8	9	10
Practically nothing									**Normal quantity**

Could you indicate the quality of the work you performed today as compared with normal on the scale below?

1	2	3	4	5	6	7	8	9	10
Very poor quality									**Normal quality**

At the end of the week the respondents had to complete the O-method and the VR-method. These questions covered the past week.

The exact phrasing in Dutch in the questionnaire was:

Hoeveel ***dagen*** *in de afgelopen week bent u* ***wél*** *naar uw werk geweest, terwijl u last had van gezondheidsproblemen?*
LET OP: dagen dat u zich ziek gemeld heeft niet meerekenen.

..... *dagen.*

Wij willen u vragen op de meetschaal hieronder te omcirkelen hoe efficiënt u volgens u gewerkt heeft op de dagen dat u ***wél*** *op uw werk was terwijl u last had van gezondheidsproblemen. Op de meetschaal betekent 10 dat uw werk niet werd beïnvloed, 1 betekent dat u zeer slecht in staat was uw werk uit te voeren.*

1	2	3	4	5	6	7	8	9	10
Maximaal Inefficiënt									**Even efficiënt als normaal**

Hoeveel uur zou u in de AFGELOPEN WEEK langer hebben moeten werken om het werk dat u niet hebt kunnen doen als gevolg van gezondheidsproblemen in te halen?
LET OP: dagen dat u zich ziek gemeld heeft niet meerekenen.

..... *uur*

Translated into English:

How many **days** during this past week did you go to work while suffering from health problems?
NOTE: Do not count the days on which you reported sick.

..... days

Please circle on the scale below the degree of efficiency with which you consider yourself to have worked with on the days you did go to work while suffering from health problems. On this scale 10 means your work was not affected and 1 means that you were hardly capable of performing your work.

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Very inefficiently									**As efficient as normal**

How many hours extra would you have had to work to catch up on tasks you were unable to complete in normal working hours due to health problems IN THE PAST WEEK?
NOTE: Do not count the days on which you reported sick.

..... *hours*

7. The valuation of informal care in economic appraisal

Summary

This chapter discusses the individual's choice to engage in informal care and the valuation of informal care in economic evaluations of health care. Besides attention for the choice between formal and informal care, also future areas of research are indicated. Traditional methods used in economic evaluation studies for valuing time spent on informal care are discussed and an alternative method is put forward. This alternative method incorporates quality of life of caregivers as an outcome-measure to represent the effects on caregivers. The methodological issues concerning the valuation of informal caregivers' time are becoming more important as new drugs and other health care technologies are introduced for patients with diseases that are typically associated with informal care.

7.1 Introduction

Informal care is a specific type of care provided by nonprofessionals. Specifically, informal caregivers are defined here as family, friends, acquaintances, or neighbors of a patient providing care for which they do not have to be financially compensated (Smith and Wright, 1994; Wright, 1987). Informal caregivers give up normal activities because of the incidence of illness in their direct social environment. An individual's choice to engage in providing informal care may not only be subject to an optimalization process of his or her utility function. Many restrictions and difficulties may influence the outcome of the mix of formal and informal care.

The valuation of informal care in cost-effectiveness analysis is rather troublesome, since available methods aim at expressing all consequences for the informal caregiver in monetary terms and incorporate the results on the cost side of economic evaluations. This practice proves rather difficult because the consequences related to informal care include effects like fatigue,

giving up leisure activities and fewer social contacts. These effects are difficult to express in monetary terms. Research seems needed in order to find methods for valuing the time involved in informal care that are theoretically correct, yet practically applicable. To this point some suggestions are made in the literature, but these do not translate easily into guidelines for practical research. For instance, capturing the full opportunity costs of time by means of a adjusted wage rate will require research on how to adjust wages adequately for people who are engaged in home production or leisure (Gold et al., 1996).

Although difficult, the valuation of time of informal caregivers becomes more important now that drugs for patients with Alzheimer's disease and schizophrenia are being developed and becoming available. These conditions are typically associated with a considerable amount of informal care. Therefore, better methods for valuing informal care are called for to ensure a good estimation of the true societal impact of informal care. Moreover, in some countries there seems to be an increasing substitution of formal care with informal care, as formal care is increasingly being restricted.

This chapter discusses an individual's choice to engage in informal care, mainly focusing on the restrictions and difficulties that potential informal caregivers face. Then, we will discuss a preferred method of valuing informal care within the context of cost-effectiveness analysis, a method aimed at finding the societal opportunity costs and effects of informal care. It is different from other proposed methods in using quality-of-life measurement for informal caregivers to avoid arbitrary monetary valuations of such aspects of informal care as fatigue, giving up leisure or sleep, etc.

The individual's decision to engage in informal caregiving is considered in section 7.2. Section 7.3 discusses the societal valuation of the inputs of informal caregivers in economic evaluations, while section 7.4 concludes this chapter and indicates areas for future research.

7.2 The individual decision to engage in informal care

Basic assumptions on individual choice

First, it is important to understand that the decision whether or not to become an informal caregiver is dependent on the incidence of illness in an

individual's direct social environment. When and if illness sets in, one option is to continue normal activities and hire professional help to care for the patient. This formal help may take place at home or may involve institutionalization. One can also perform some or all caregiving activities oneself and give up activities that one was normally engaged in. Often, the decision is not all or nothing but rather a decision to combine both formal and informal care.

Second, it may be argued that when a relative or friend falls ill and is in need of care, both formal and informal care yield utility for the potential informal caregiver. The alternative is to leave the patient with insufficient care, or in the worst case without care which will be associated with a large disutility. As the wellbeing of the patient is assumed to be incorporated in the utility function of a relative or friend, any improvement in the patient's condition will have a positive impact on his or her utility. If, for instance, a patient has a very strong preference to stay at home, the disutility from institutionalization will be relatively high, both for the patient and the potential informal caregiver. The interaction between utility of patients and informal caregivers is an interesting and challenging subject. The amount of utility from informal care and formal care is bound to differ both for the patient and the informal caregiver, since both involve a different time input of the caregiver, different financial costs, and a different process utility for the caregiver.

Third, the choice between formal and informal care as discussed here is relevant for only that part of formal care that may be substituted by informal care. It is obvious that some care activities are too complex to be performed by a nonprofessional.

Finally, a difference can be observed between members of the same household acting as an informal caregiver and someone else acting as an informal caregiver. Some of the costs associated with informal and formal care will have to be paid by the patient and thus may be expected to be of interest to members of the same household in case of a shared budget. Other informal caregivers may consider only the costs and effects that they will incur themselves. These two groups will not be consequently disentangled in the following discussion, but it is assumed to be obvious for the reader when to make this distinction.

The choice between formal and informal care

If a person decides to become an informal caregiver, it is clear from a revealed preference viewpoint that *given the circumstances* this is the utility maximizing solution, because otherwise he would have chosen differently. This means that if informal care is provided, the positive and negative sides of informal care together yield more utility than the positive and negative aspects of the alternative, formal care and both yield more utility than no care at all. This will be the starting point for the discussion, but later we will concentrate on the fact that real world circumstances often limit opportunities. From the starting point of the incidence of illness onward both formal and informal care yield utility, because the alternative is no care at all. Formal and informal care may then be viewed as two distinct types of care by both patients and caregivers. Both types of care and any combination of them are associated with different time input, financial costs, and process utility (providing care). An individual is supposed to be influenced by these aspects of informal and formal care when striving towards an optimal formal-informal care mix. Without trying to be complete, some of these positive and negative sides of formal and informal care are discussed below.

Informal care may involve hiring or buying certain attributes, such as buying special beds and adapting one's home, in order to facilitate the provision of informal care (Glendinning, 1992). Travel costs may be important when an informal caregiver does not live nearby. Hiring formal care may be costly as well, since the related costs may often not be entirely refunded by social or private insurance. Travel costs can also be of importance for formal care, for instance when visiting an institutionalized patient.

A major input in the treatment of the patient of an informal caregiver is time. The time allocated to informal care may otherwise have been invested in paid or unpaid work and/or leisure. The opportunity costs of time invested in informal care are equal to the valuation of alternative activities during that time, because it is the time use that is valued by persons, not time as such (Layard and Walters, 1978). The opportunity costs of this time-input are expected to differ between formal and informal care, since the amount of time invested by informal caregivers in either case is bound to differ. The valuation of this time-input in a cost-effectiveness analysis is discussed later.

Providing informal care may affect the wellbeing of the informal caregiver since it involves all kinds of pleasant and unpleasant activities. Cleaning up after the patient has been sick, waking up in the middle of the night because the patient needs attention etc., may have negative effects on the caregiver's utility. Positive effects are derived from being close to the patient, being able to take care of him or her, enjoying certain caregiving activities etc. These positive and negative sides to informal care have been discussed extensively by Wright (1987). These effects are closely related to the caregiving activities that replace normal activities, as we will discuss in section 7.3.

There will also be direct effects of formal care on the wellbeing of the relative or friend. When institutionalization is the alternative of informal care, the absence of the patient may have great impact on the wellbeing of the potential caregiver. This impact becomes even greater when the patient prefers to be cared for at home, since the wellbeing of the patient is assumed to be incorporated in the utility function of the relative or friend. When formal care can be provided in the patient's home, and the caregiver is part of the same household, considerations of privacy and discomfort of having a strange caregiver in the house may also influence the exact mix of formal and informal care.

Individual choice under ideal circumstances

In choosing between formal and informal care, the numerous aspects mentioned above may all enter the utility function of the potential caregiver and may all be considered to influence the final decision. The approach chosen here is to compare the utility from provision of informal care as such to that of hiring formal care. The total costs and benefits attached to formal and informal care may be considered. By taking into account all costs and effects of both types of care, the optimal mix of formal and informal care may be found.

In the most simple case, formal and informal care may be seen as perfect substitutes. Then, for a potential informal caregiver with paid work, in a Gronau household production model the optimal spending of time may be denoted as:

(1) $U = U(Z_i + Z_f, Z_c, Z_l)$

in which U represents the utility of the potential informal caregiver, which is dependent on four endproducts: formal care (Z_f) purchased for one's partner, informal care (Z_i), leisure (Z_l) and other consumption (Z_c). Note that the formal and informal care is used to care for the partner or friend of this person. The substitutability of formal and informal care is formulated by making utility dependent on the sum of Z_i and Z_f.

Further, we will assume informal care (Z_i) to be equal to time spent on informal care, t_i, thus:

(2) $Z_i = t_i$,

formal care (Z_f) is equal to the amount of formal care, F, puchased:

(3) $Z_f = F$

leisure (Z_l) is equal to all time not spent on paid work or informal care (t_l):

(4) $Z_l = t_l$

and other consumption (Z_c) is equal to the amount of other goods, C, purchased:

(5) $Z_c = C$

All the available funds, partly earned by performing paid labor and partly other (non-labor) income (Y_0), will be spent on formal care and other consumption, so the budget restriction is:

(6) $p_f F + p_c C = t_w W + Y_0$

in which p_f is the price of formal care per unit, p_c is the price per unit of other consumption goods, t_w is the amount of hours spent on paid work and W is the wage rate. Furthermore, the total amount of time is restricted:

(7) $T = t_l + t_i + t_w$

Equations (6) and (7) may be rewritten to the total income equation:

(8) $WT = p_f F + p_c C + Wt_l + Wt_i$

Using a Langrange function, optimal time spending holds when:

(9) $\text{Max } Z = U(F + t_i, C, t_l) + \mu[WT - p_f F - p_c C - Wt_i - Wt_l]$

From this one may derive the substitution rate between informal and formal care:

$$(10)\ \frac{\partial Z}{\partial t_i} = \frac{\partial U}{\partial t_i} + \mu W = 0$$

$$(11)\ \frac{\partial Z}{\partial F} = \frac{\partial U}{\partial F} + \mu p_f = 0$$

Combining (10) and (11) leads to

$$(12)\ \frac{\frac{\partial U}{\partial t_i}}{\frac{\partial U}{\partial F}} = \frac{W}{p_f}$$

Marginal substitution rate between providing informal care and purchasing formal care should therefore be equal to the price relation.

Of course, even in a textbook situation, the point at which one substitutes informal care by formal care will differ among persons and illnesses. Informal care may almost always yield more utility for some persons, while for others hiring formal care will be preferable at almost every point. Especially when the potential caregiver has a paid job, it may be difficult to be absent for a longer period, and thus formal care may be preferred relatively soon. Note that if a higher educated (and higher paid) person provides substantial amounts of informal care, revealed preference may indicate that the productivity of this person is high not only in the marketplace, but also at home. Empirical studies on the choice of informal caregivers are very scarce.

The amount of care a patient needs per day or week is of great importance. The utility derived from formal and informal care may shift over time. After having given informal care for several months, an informal caregiver might value formal care higher as compared with when he had just begun providing informal care (burn-out). This time effect on the formal-informal care decision would be extremely interesting to explore further in empirical work as well.

Real life restrictions on optimal choice

The rational and 'unrestricted' optimalization behavior, as described above, will not be found in everyday circumstances. Here, we consider the importance of real-life restrictions on the optimal choice as described above. One complicating factor in real life is uncertainty. When deciding between formal and informal care, it may be unclear how long care is needed and thus the decision whether or not to engage in informal care is made under uncertainty. Such a decision under uncertainty may perhaps be pictured better through a simple decision tree in which the potential informal caregiver is unaware of the length of the care period. In figure 7.1 an example of a decision tree (or game tree) is provided for an informal caregiver. We will assume that the person in this example would prefer to provide informal care in case of a short period of illness and to hire formal care in case of a longer period of illness.

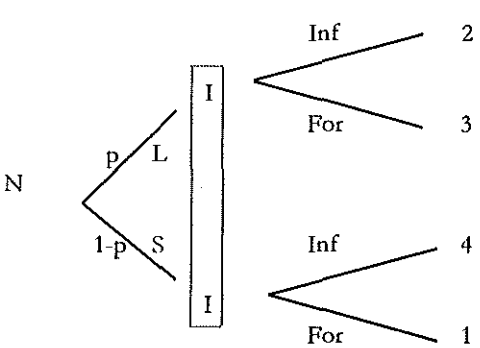

Figure 7.1: Example of a decision tree for an informal caregiver.

This decision tree has subjective, exemplary values that show the utility pay-off of each choice of the potential caregiver. The individual I has to decide

whether or not to give informal care, without knowing whether the patient will be ill for a long (L) or a short period (S), which is decided by nature (N). The dotted lines indicate that the individual is unaware of his or her position in the decision tree. A lengthy illness has a chance of p, a short one a chance of 1-p. Depending on the length of the illness informal care or formal care may be the best solution. A lengthy illness might cause considerable distress for the informal caregiver (for instance reduced performance on paid work). However, a short period of informal care may be preferred over a short period of formal care, because of relational aspects, for instance. Under such conditions a person again will try to consider all costs and effects, and in this example the informal caregiver will perform informal care if he *believes* that the chance of long illness is less than 0.75[16]. Such decisions under uncertainty may occur frequently; therefore, observed behavior does not have to be the result of optimalization with perfect information, but based on *beliefs* of the informal caregiver.

Another problem is the substitution of formal care by informal care. Earlier we made the assumption that informal caregiving can be substituted with formal care at any point. However, this may not always be possible in reality and the decision sometimes is more or less a 'to care or not to care' decision (Wright, 1987). For instance, if the only reasonable alternative for extensive informal care involves institutionalization of a patient, the informal caregiver has to decide which alternative yields the highest utility or the lowest disutility. In general, discrete or 'lumpy' decisions about the number of hours spent on informal care and/or formal care are common and may be modeled in a more formal analysis, which goes beyond the scope of this chapter.

It may also be the case that only a limited amount of formal care can be hired, when insurance does not cover certain types of formal care and the household budget does not allow for it to be paid out of pocket. The choice for informal care can also be a result from, for the informal caregiver exogenous, supply restrictions for formal care. When budgets are cut by governments, formal care may be rationed in order to provide a maximum amount of people with a minimum of care per patient. The remaining need for care will then have to

[16] Expected utility from choosing informal care: $2p + 4(1-p)$
Expected utility from choosing formal care: $3p + (1-p)$
Indifference when $2p + 4 - 4p = 3p + 1 - p$
$\Rightarrow -2p + 4 = 2p + 1$
$\Rightarrow -4p = -3$ thus $p = 0.75$

be hired privately (if possible), or informal care has to be provided from that point onward.

Informal caregivers may not always be able to provide all the care they would want to. In that case, informal and formal care are complementary rather than substitutes. For instance, if an elderly woman takes care of her sick husband, she may be unable to perform physically demanding caregiving tasks, such as bathing her husband. This kind of caregiving task may then be performed by a formal caregiver, while she herself engages in the remaining caregiving activities.

Finally, the preferences of the patient can influence the decision of the caregiver. How does the patient's utility interact with the caregiver's utility? Often it is just assumed that patients' utility is incorporated in the utility function of the informal caregiver (which naturally may be correct); however, it would be interesting to see exactly how these two quantities interact. It is recommended here that all of the aspects mentioned above be investigated further empirically, since little empirical work on many of these issues is available. In Table 7.1 the main arguments entering the caregivers utility function and important restrictions are summarized.

	Costs and Health/Welfare Effects
Formal Care:	Costs: Travel costs if patient is institutionalized, costs of hiring formal care (if not totally refunded),time costs Effects: Privacy, absence of patient in case of institutionalisation, patient's utility (preferences)
Informal Care:	Costs: Travel costs, costs of home-adaptations and consumer durables, time costs (giving up paid work, unpaid work and/or leisure) Effects: Performing tasks that are unpleasant or at awkward times (e.g. at night), fatigue, patient's utility (preferences)
	Important Restrictions
Formal Care:	uncertainty (e.g. about length of illness), budget restrictions, external restrictions on amount of formal care, sometimes formal-informal mix not possible, e.g. when institutionalisation is only alternative.
Informal Care:	uncertainty (e.g. about length of illness, progression in care-intensity), physical restrictions to perform all caregiving tasks, sometimes formal-informal mix not possible.

Table 7.1: Summary individual choice

7.3 Societal valuation of time invested in informal care

The valuation of time input of informal caregivers is rather difficult, especially since some of the activities performed are associated with intangible effects like fatigue. The valuation of these effects will also be discussed in this section. First, we will discuss well-known valuation methods of time of informal caregivers. Then we will present an alternative valuation method of time invested in informal care. Finally, the incorporation of the results in a cost-effectiveness analysis is discussed.

Time: amount and valuation
Time spent on giving informal care is a resource invested by family and friends and should therefore be valued in an economic evaluation. Two main problems can be detected when trying to incorporate time costs into CEA: the amount and the value of time. The first problem is assessing the exact quantity of time spent on informal caregiving. During the time providing informal care, many normal activities often can continue as usual. Thus joint-production occurs, as in the case of surveillance (Busschbach et al., 1998). It may also prove to be difficult to separate normal activities from caregiving activities. Preparing and serving a meal is often considered a normal task; however, helping someone to eat, in most cases is not. These problems may be solved by using a structured interview or questionnaire asking specifically about loss of normal time use on paid work, unpaid work and leisure (Busschbach et al., 1998).

The second problem is the valuation of the time input into informal care. Several approaches are possible: the market price method, the opportunity costs method and contingent valuation.

In the market price method, the time input of informal caregivers is valued at its market price. This market price is mostly equal to the costs of hiring a professional caregiver to perform the caregiving activities (Rutten et al., 1993). This method can be seen as a valuation method of output rather than of input (Posnett and Jan, 1996), since the output of informal care is valued at a price equal to that of formal care. The most important objection against valuing informal care with this practical method is that the value that is ultimately attached to informal care may have little or no relation to the amount and value of the inputs sacrified in informal care (the opportunity costs of inputs).

The second way to value informal care is to look for the opportunity costs of the time spent on informal care, which involves the valuation of input (Posnett and Jan, 1996). These opportunity costs are set equal to the value of the best alternative use of time by the informal caregiver. The best alternative use is assumed here to be the normal use of time, following a revealed preference viewpoint (which is questionable, but alternatives are not easy to find). This opportunity costs approach is often recommended to be used in economic evaluations (Gold et al., 1996). Generally, this valuation method considers lost working time by taking the wage rate. Although this valuation method is widely accepted, it should be noticed that it ignores any nonmonetary utility that may be derived from working. The valuation of time otherwise spent on leisure activities or unpaid work is more difficult (Gold et al., 1996; Smith and Wright, 1994). One method to value unpaid working time is to use the wage rate that a person receives when working for pay, or a similar person receives as wage (Gold et al., 1996; Posnett and Jan, 1996). Wage rates serve here as a lower bound estimate of full opportunity costs (Gold et al., 1996). The reasoning behind this is that if full opportunity costs would be below the level of the wage rate, the person would have worked for pay instead of engaging in such unpaid activities as home production. However, exactly how the wage rate relates to the gain from unpaid activities is unclear. Although it is sometimes assumed to be 'close enough to the real opportunity costs' (Gold et al., 1996), the empirical evidence for this assumption is lacking, especially for people who are restricted in their choice by the labor market situation or by personal, institutional or other factors.

Leisure time is sometimes completely ignored in economic evaluations. Although appealing, adjustment of wage rates to serve as a proxy for opportunity costs of leisure time is rather difficult. In transport appraisals leisure time was valued at 40% and later at 24.7% of working time (Smith and Wright, 1994). One may argue that this seems an underestimation of the value of leisure time because a substantial proportion of the work force is engaged in part time work, trading paid work for leisure. Also, if one bases the valuation on microeconomic models of time allocation, as Posnett and Jan (1996) elegantly do, one would expect a valuation of leisure time equal to (or higher than) the marginal valuation of working time.

Although the opportunity cost method is theoretically the most appealing, its translation into practical guidelines necessarily seems to involve somewhat

arbitrary and difficult decisions to find monetary values for time uses that cannot be expressed in monetary terms easily. An opportunity cost-based method is sought that would try to avoid an arbitrary monetary valuation of unpaid working time and leisure time, yet would give practical and applicable solutions.

Another possible option is contingent valuation. In such a method, one can ask informal caregivers or the general public to assess their willingness to pay for no longer having to perform informal care activities or their willingness to accept having to perform them. This would provide an overall estimation of the costs of informal care, but theoretically only time aspects can be focused on as well. Although this method is appealing, there are concerns about the validity and consistency of the valuations provided. Apart from concerns about the operationalization of such methods (information dissemination, type of question used and framing bias) there are also concerns on the validity of the actual answer, even if operationalization were perfect. Would the willingness-to-pay amount actually be paid if the situation described occurred? There may be strategic behaviour by the respondents, but also one might question whether people are capable of validly valuing something as complex as informal care.

If we look at two important publications on the methodology of cost-effectiveness analysis, i.e., that of Gold et al. (1996) and that of Drummond et al. (1997), both publications offer little practical guidance as to how informal care should be valued. Drummond and colleagues (1997, page 86) mention that 'in general the valuation of this raises the same issues as the valuation of patients' time' and they recommend for patient time the use of gross salary when absent from work and that 'different assumptions can be made about the opportunity costs of leisure time'. These recommendations are very similar to those of the U.S. Panel on Cost-Effectiveness in Health and Medicine. They state that the time of informal caregivers can be valued by using 'its market or reservation price, similar to the methods suggested for valuing time of housewives and househusbands'. (Gold et al., 1996, page 203) However, neither offers a method of adjusting wage rates for time spent on unpaid work or leisure.

Valuing time spent on informal care: an alternative approach

Here we will present an approach that was initially aimed at establishing time costs of patients, however, the approach can also be used to value time of

informal caregivers (see chapter 3). In three brief sections, valuation of paid work time, unpaid work time and leisure are discussed.

Paid work: When the caregiver gives up paid labor time, this time input can be valued with a method like the friction cost method (Koopmanschap et al., 1995; Johannesson and Karlsson, 1997; Koopmanschap et al., 1997). In the context of this chapter a thorough discussion of the friction cost method is not feasible. The friction cost method measures production losses due to absence from a societal viewpoint. It recognizes possibilities to replace a person who gives up paid work time in order to provide informal care. Lost utility because of lost role functioning in a paid job may be captured in terms of quality-of-life. Incorporating lost income in quality of life measurement is not considered a valid method, since the relationship between productivity, income and quality of life is quite arbitrary (Brouwer et al., 1997a; Weinstein et al., 1997; Brouwer et al., 1997b).

Unpaid work: Unpaid work is seen as yielding two types of utility: process utility and product utility. Product utility is utility gained from the output of home production, the produced items themselves (as may be produced by a professional housekeeper). The productive value (product utility) of home production may best be estimated by taking the wage rate of a professional housekeeper. Future research may be aimed at investigating the possibility of replacement of unpaid work. Process utility is the utility gained from the process of performing the tasks yourself. Especially since home production may involve aspects such as raising children, the 'production-process' may in itself yield a considerable amount of utility. A good estimation of the opportunity costs of this aspect of unpaid work seems essential. In our opinion this utility may best be measured in terms of quality of life. This should ensure a true estimation of the impact involved, and avoid the use of an arbitrary monetary estimation of the opportunity costs involved. In for instance the EuroQol/EQ-5D, a standard classification instrument, this aspect already has been captured (for patient time), albeit not in the most sophisticated manner perhaps. This shows that there seems to be some consensus that this part of home production is related to 'home role functioning' and therefore should be captured in quality of life.

Leisure time: Leisure time contributes positively to one's quality of life and may be valued highly by persons engaged in informal care. Leisure time may relatively often be sacrificed when elderly retired persons are involved in

caregiving (Wright, 1987). However, to capture this loss in quality of life in a monetary amount seems difficult, if not impossible. Therefore, we propose to value lost leisure time through quality of life and not monetarily. Again, in many quality-of-life instruments (e.g. EuroQol/EQ-5D), the ability to engage in leisure activities is already included as an aspect of quality of life. Johannesson (1997) also feels that it 'seems reasonable to assume that individuals take into account the change in leisure in assessments of quality weights...'

Quality-of-life effects of caregiving activities: When using a quality-of-life questionnaire, the quality of life of informal caregivers may be assessed while also considering the impact on quality of life of the caregiving activities they perform. It seems useful to incorporate these effects of informal care on the informal caregiver in the analysis in terms of quality of life. Other valuation methods are dependent on an often arbitrary, monetary valuation of these 'intangible effects' such as the disutility from having to perform activities that are unpleasant. It would be practical to combine all effects on quality of life in one measurement method, i.e., one questionnaire.

For instance Drummond et al. (1991) and Mohide et al. (1988) have already performed some work in this area. Drummond and colleagues evaluated a support program for caregivers of demented elderly. The fact that relatively large changes in the quality of life of caregivers were found in their study provokes the question of why caregiver quality of life is not measured as a normal part of CEA when informal care is significantly present. This would give explicit attention to significant quality of life changes of caregivers rather than making them a part of total costs. The above described method would make such a shift from costs of informal care to effects on caregivers possible without double-counting. Mohide et al. (1988) used the time trade-off technique to assess well being of family caregivers. They found that this method, normally used to assess patient quality of life, appeared to be feasible, reliable, valid, and responsive to change. Therefore, the quality of life of caregivers seems to change due to caregiving *and* it seems possible to measure this change in a valid way.

Incorporation of results in the analysis

The change in quality of life of caregivers due to the caregiving activities should in principle be measured in an economic evaluation such as a cost-effectiveness analysis. One might argue that the incorporation of informal

caregivers' effects in an analysis in terms of quality of life is closer to the concept of taking the societal perspective than to 'force' these effects into the cost-side of the analysis. The U.S. Panel on Cost-Effectiveness in Health and Medicine also mentions that the 'societal perspective prescribes that consequences for all affected persons should be included in the analysis' and encourages analysts 'to think broadly about the people affected by the intervention and begin to include health related quality of life effects of significant others in sensitivity analyses when they are important' (Gold et al., 1996, page 67). Of course, the U.S. panel seems to refer to 'family effects', i.e. effects of the outcomes of an intervention on family and friends, rather than to the effects on family and friends providing inputs in the intervention itself. However, since the efforts of informal caregivers can be seen as inputs to the intervention itself, it seems even more appropriate to incorporate their quality of life changes in the analysis. Nonetheless, the distinction between effects from such input and those from the disease or the intervention should be noted. Here, only the main focus is on effects from providing informal care. Parker for instance states:

'*I have tried to raise in readers' minds some doubts about the advisability of measuring the quality of life of any single individual, without also taking into account the quality of life of those who support and care for him or her.*' (Parker, 1990, page 127)

Parker seems to imply that she wants to measure both 'family effects' and informal care-effects in terms of quality of life, which is a broader view of caregiver quality of life than the one discussed here. Similarly, Baldwin and Gerard (1990, page 145) indicate that in families caring for mentally handicapped children the costs attached to this care

'*...are significant and diverse, potentially affecting the quality of life of all members of the family.*'

And
'*in order fully to evaluate the effects of respite care it is necessary to face up to the difficult task of trading off gains and losses in the quality of life of different family members*'. (page 146)

In some instances, such as informal care for a patient with Alzheimer's disease patient who already suffers from significant cognitive deterioration, one might consider the caregivers' quality of life to be just as or even more important as

that of the patient perhaps. Incorporating these effects in terms of quality of life ensures that the impact on caregivers will receive explicit attention. By incorporating these effects in monetary terms, this attention will probably be less explicit and the estimations of the impact may be less accurate. From a societal perspective, it is unclear why a quality of life change of a caregiver should have less explicit attention than that of a patient. As mentioned earlier, caregivers' quality of life has been measured in earlier studies and used for a cost-effectiveness analysis (Drummond et al., 1991), but according to our knowledge never combined with patient quality of life. This should not be surprising, since the denominator of the cost-effectiveness ratio is exclusively reserved for *health-related* quality of life, whereas the caregivers' quality of life-change can be considered to be more *care-related.*

This issue is somewhat dependent on the questionnaire used. For instance, if only a EuroQol/EQ-5D questionnaire were used, with the explicit notion that the respondent should incorporate changes due to providing informal care, one might conclude that only health-related quality of life changes are measured, since that is the aim of the EuroQol/EQ-5D instrument. Influences on the mental and physical health of the caregiver and on his or her social activities may then be considered as changes in health-related quality of life and balanced with that of the patient. However, this health-related quality of life change of the caregiver is not normally incorporated in a cost-effectiveness analysis.

On the other hand, by specifically considering informal care activities one might argue that the quality of life impact registered then is care-related rather than health-related. The obvious advantage is that such instruments may be more sensitive. To combine patient and caregiver effects in the cost-effectiveness ratio when quality of life measurement is more care-related will ultimately lead to optimalization of health-related and care-related quality of life. Though this may have certain advantages, it is unclear whether such a change is to be considered an improvement or a degradation of cost-effectiveness analysis. To use multicriteria analysis for studies in which informal care plays a significant role seems a compromise between ignoring caregiver effects and 'just' combining them with health related patient quality of life. Also, using the pro memory informal caregiver quality of life change in the sensitivity analysis as suggested by the U.S. Panel may be considered a useful method of incorporating the results in the analysis. Aggregation of caregiver and patient quality of life, especially when measured with different

instruments, of course raises numerous questions of how this should be operationalized in a valid manner.

At this point, measuring informal caregivers' quality of life and developing and validating questionnaires to do so, is recommended here. It seems an appropriate tool to use in studies in which informal care plays an important role. Future research may also be aimed at finding ways to balance caregiver and patient quality of life.

7.4 Discussion

More empirical research on the individual's decision to engage in informal care is needed to increase our knowledge on this subject. In such research, not only actual informal caregivers should be interviewed, but also potential informal caregivers who have decided to hire formal care. In this way we can attempt to estimate the utility functions underlying their decisions, taking into account real-life restrictions and perhaps design support programs that would reduce the strain on informal caregivers. This would ensure a maximum use of the possibilities for substituting expensive institutional care for home care, which may reduce costs and improve the quality of life both for informal caregivers and patients.

Volunteers, people engaged in caregiving in their leisure time, form a specific group of informal caregivers. They have incorporated caregiving willingly and freely into their normal activities. Therefore, in a revealed preference framework one may argue that they do not give up a better activity for caregiving from their individual perspective, but choose to perform care tasks as the best possible use for part of their leisure time. In that respect, no *individual* opportunity costs from caregiving may be incurred. However, one might claim that volunteers who would not care for these patients would free resources, i.e. time, for other alternative uses (like caring for other patients), and thus volunteer care does lead to opportunity costs from a societal perspective. Which of these approaches is considered most appropriate remains unclear.

The societal valuation of informal care proves a difficult subject. The alternative method presented here is believed to be a first step towards a reliable estimation of the total opportunity costs involved. This method

should be developed further by constructing internationally applicable questionnaires or techniques in which all aspects mentioned can be incorporated. Existing questionnaires, such as the *Caregiver Quality of Life Instrument* used by Drummond et al. (1991), the Time Trade Off-technique used by Mohide et al. (1988), the *Caregiver Hassle Scale* (Kinney and Stephens, 1989) and *The Burden Interview* (Zarit et al., 1980), should be examined for their appropriateness to use together with the costing principles discussed. Further research may also be conducted in order to investigate the desirability to incorporate caregivers' quality of life into the effect side of the cost-effectiveness ratio. Until better solutions are available, a multi criteria approach or incorporating results in the sensitivity analysis seems most appropriate for studies in which informal care is important. Especially now that drugs are becoming available for patients with diseases often associated with a considerable amount of informal care, such as Alzheimer's disease and schizophrenia, the valuation of time of informal caregivers becomes more important. Therefore, it seems useful to further investigate the proposal discussed in this chapter in order to ensure a good estimation of the true societal impact of informal care.

8. Priority care for employees: a blessing in disguise?

Summary

This chapter discusses the efficiency and equity effects of priority care for employees. Recent privatization of workers' compensation insurance in the Netherlands caused an increasing tension between public responsibility for health care cost containment and private responsibility for sick pay. As a result of strict supply side regulation, waiting lists increased while at the same time employers became fully responsible for sick pay. To reduce sick pay and production losses, employers are prepared to pay for priority care by using available excess capacity. We argue that the criteria of Pareto and Rawls can provide a rationale for the resulting differential treatment of employees and non-employees. However, such a justification crucially depends on weights society assigns to absolute versus relative improvements in access to health care.

8.1 Introduction

In many countries waiting lists in health care are an important issue, because cost-containment and regulation result in rationing of health services. Although waiting lists may have a use in terms of selection and prioritization, waiting for an obviously necessary intervention is often perceived as undesirable. Loss of quality of life due to waiting, having to undergo additional treatment and deterioration of health while waiting are all arguments to advocate a minimal waiting time. Many countries have experimented to reduce waiting lists without relaxing the control of overall (public) expenditures, but many of them failed (Hanning, 1996; Newton et al., 1995).

A solution to reduce waiting time which may be considered is to attract extra funding for health care, specifically aimed at treating certain groups of

patients with priority. In the Netherlands this was operationalised by starting initiatives to treat employees of certain firms (who were willing to pay for this) more swiftly than normal. By treating employees outside normal working hours of hospitals, waiting lists were avoided and a quick recovery and return to work was assured.

Important to notice is that in the Netherlands uniform treatment of all (regardless of socio-economic status) is a broadly accepted principle of equity in health care. Therefore, initiatives that seem to undermine this principle are bound to receive a lot of criticism on equity grounds. In other countries with a largely publicly financed health care system the egalitarian principle is less strictly applied. For instance in the UK the differences in treatment (especially in waiting time) between privately insured and patients fully dependent on the National Health Service (NHS) is quite common. At least at the practical level, a two-tier system is accepted there, perhaps partly to decrease the pressure on the NHS by providing some with other sources of care. Since the UK-system is largely tax-funded, people treated in private hospitals have to pay their full share to the NHS and on top of that the costs of private treatment (or an insurance covering private insurance). In this way at least equity on the finance-side is assured.

In this chapter we will discuss the ongoing debate in the Netherlands, investigating the motives for priority care for employees and the responses from different organizations. Also, we will investigate whether giving priority in care on other grounds than medical need may be justified under certain conditions. Finally, we will discuss whether or not priority care for employees might constitute a blessing in disguise.

8.2 The Dutch discussion

The Dutch government has focused a lot of attention on reducing absence from work due to illness in the early 90's. The aim of the government was to decentralize the responsibility of sickpay in order to provide companies with direct financial incentives to reduce absence as much as possible. In two steps the government introduced a radical privatization of social workers' compensation insurance. In 1994 the responsibility to continue

payment during the first 2 or 6 weeks of absence (depending on the size of a firm) was transferred from social insurance to individual companies, by introducing a law which legally binds firms to pay sick employees (the WULBZ). This change had potentially major financial consequences for individual firms, especially for smaller ones. Therefore, it was not surprising that many investigations showed a significant decrease in absence rates between 1993 and 1994, from about 7% to about 5,5% (CTSV, 1995a). Hence, the transfer of financial risk to the individual firms had an important impact on absence. In 1996 the individual firm's responsibility for sick pay was expanded from 2 or 6 weeks to one year. Again, the Dutch Central Bureau of Statistics found a decrease in absence rates, although less dramatic, from 5.5% in 1995 to 5.1% in 1996 (CBS, 1997).

In search to further reduce absence from work, firms stumbled upon waiting time for medical care. In the Netherlands waiting for treatment by a specialist is quite common. To give an indication of average waiting time for different specialties, both for inpatient and outpatient care, Table 8.1 presents the results from the Health Interview Survey 1993/1995 performed by the Dutch CBS (Swinkels, 1996).

Specialist	*Average waiting time in days*	*Acceptable maximum waiting time (indicated by patients)*
Gynaecologist	51.8	41.7
Orthopedist	66.5	23.0
Eye doctor	88.5	30.6
Neurologist	40.1	11.2
Psychiatrist	57.4	23.4
ENT-specialist	50.6	28.9
General surgeon	59.6	23.8
Plastic surgeon	140.1	55.5
Total	**58.3**	**24.2**

Table 8.1: Waiting time in the Netherlands; real and acceptable

As is shown in Table 8.1, on average patients had to wait 58 days before they received proper treatment. In Dutch society this is widely considered a

long time and patients in a survey indicate that the waiting period exceeds the acceptable maximum waiting time by a factor 2.5 on average for all these specialties.

The excessive waiting periods cause a tension between the private responsibility of firms for continuation of sick pay and the public responsibility of providing health care for all. Although the Dutch government chose to privatize social workers' compensation insurance, the health care sector is still very much centrally organized. With macro-budgets for health care, supply is restricted in order to control expenses. Hospital supply is restricted by capacity regulation, by regulation of the number and type of medical specialists and by annual budgets for hospitals' operating expenses. One of the consequences of these regulations are waiting lists; hospitals cannot provide everybody swiftly with the appropriate care because their budget constraints do not allow them to operate at full capacity. At present there is substantial amount of unused capacity in the hospital sector.

The increasing tension between private and public reponsibilities for health care costs brought on a renewed discussion about waiting lists. Some employers, insurers and hospitals have joined forces to create a so-called 'employee-clinics'. These clinics are aimed at providing priority care for employees in order to facilitate a swift return to work. Care provided by these clinics is produced by using formerly unused hospital capacity (weekends, evenings) and paid for directly by the employers. In first instance, the Dutch Minister of Health, Minister Borst, seemed to cautiously support these activities in the short run, although she announced that a more permanent viewpoint would follow. Surprisingly, the permanent government stance turned out to be a total rejection of the initiatives to start employee-clinics. The Minister stated that priority care for employees violates the principle of equal treatment for all and that prioritization should be based solely on medical need. The debate concentrated on the position of non-employees. Opponents of employee-clinics claimed that it would be unethical for non-employees to have to wait longer than employees and that this group would be disadvantaged by employee-clinics. The majority of the political parties also opposed explicit priority care for employees, supporting the government's stance that any difference in medical treatment between employees and non-

employees is unacceptable. To alleviate the pressure from firms, unions, employer organizations and some hospitals in favor of introducing employee-clinics, the government decided to invest 50 million guilders to reduce waiting time in 1997. However, examples from other countries suggest that such investments yield results that are rather disappointing (Hanning, 1996; Newton et al., 1995). Awarding hospitals who have long waiting lists may even constitute a perverse incentive, creating rather than solving waiting lists. For example, the director of a large general hospital frankly admitted that medical specialists in his hospital approached him with the question whether they should increase their waiting lists in order to capture a part of the extra budget of 50 million for the reduction of waiting time.

To reduce waiting time for all, and to relieve some of the pressure to introduce employee-clinics, other solutions were put forward, such as the introduction of uniform categorisation of (severity of) diseases and a better provision of information about waiting lists and waiting time in different hospitals, to GP's and patients. Uniform categorisation of diseases across hospitals and specialists should ensure that waiting lists in different settings for a specific treatment may be compared. This could enhance the possibilities of treating the most severe patients first. Better public availability of information on waiting lists in different hospitals may enhance swifter treatment. At present, it may occur that in one hospital the operating room is empty, while in another people are queuing to 'get in'. Although these measures may be useful in reducing the variance in waiting time, it is questionable whether they will substantially reduce average waiting time.

As long as waiting time remains excessive, it is obvious that cost conscious firms will try to avoid to have employees on waiting lists. This pressure will lead to illegal or unwanted priority care for employees within regular hours, thereby increasing waiting time for others. Also, the increasing possibilities to be treated in other countries may be used and private clinics (which the Minister of Health effectively wants to ban as well) may focus on quick care for employees. In the latter cases, the public health care sector does not benefit from the additional funds.

8.3 Who are affected by employee-clinics?

In this section we will discuss the consequences of employee-clinics for all affected parties.

Employees
For employees the introduction of employee-clinics will mean that if their employer is willing to 'join' an employee-clinic, they will receive care more swiftly than before. They can use the care facilities their employer has purchased. Since the care will be provided by professionals, in a normal medical facility, the quality of priority care is expected to be similar to regular care. Since most will prefer swift care over slow care, i.e. a short waiting time over a longer one, it seems safe to say that employees will benefit from the introduction of employee-clinics. Of course, a necessary condition is that the employee will have to keep a final say in how he or she wishes to be treated, together with his or her general practitioner. This condition will ensure that although employers and company physicians may press for quicker care (which in many cases will have the preference of the patient as well), the employee must be in control over the type of care he receives and prefers in specific cases.

Employers
Since employers bear the additional costs of replacing a sick worker or of production losses, they will try to limit these costs. If the costs of priority care in an employee-clinic is less than the costs related to the absence of the sick workers an employer will probably be interested in paying for priority care. This will reduce his total costs by ensuring a swift return to work. In fact, the reduction of costs is an important incentive for an employer (at least a profit maximizing one) to consider paying money for a swift treatment of one of his employees. That the potential of cost reduction is large, may be illustrated by a calculation by the Netherlands Hospital Institute (NZi, 1997). They estimated that an investment in reducing waiting time of 175 million Dutch guilders would induce a savings in terms of sick pay of 1 billion Dutch guilders. This means that an improvement of the employer's position is not only possible by introducing employee-clinics, but it seems an important basis for employers to be interested in employee-clinics.

For individual firms, the trade-off is easily calculated. Consider a typical employee, producing an added value of 60,000 Dutch guilders annually and working 220 days a year, five days a week. Per day this means an average production value of about 275 guilders. If this employee has to wait for a treatment by an orthopedist (66.5 days on average; see Table 8.1) this would mean an estimated loss of 5/7*66.5 = 47.5 working days. The estimation of total production loss would be 47.5*275 = 13,062.50 Dutch guilders. Obviously, by buying formerly unused capacity in a hospital to avoid excessive waiting, huge savings can be established, even when employers bear the full costs of priority care.

Note that this example does not consider possible compensation mechanisms, like colleagues taking over some of the tasks of the absent employee. This may cause production losses to be lower than calculated here, especially for short absence spells. However, such compensating mechanisms (working overtime or having an internal labor reserve) are usually not without costs either.

Hospitals
For hospitals employee-clinics will offer the opportunity to differentiate and create income beyond their strict budgets that are determined by the government. Also, they will be able to use capacity that has been left unused up till now. If the Dutch government would allow these hospitals and other care-organizations to keep the money they earn this way, this will also mean an improvement of the position of hospitals. Moreover, a good registration of waiting time should allow hospitals to prove the reduction in waiting time, not only of employees but also of non-employees. If hospitals fail to demonstrate the improvement of the situation of both parties, government may prohibit acquiring additional funds from priority care.

Non-employees
The position of non-employees is the most controversial. Many authors and organizations have claimed that they would be disadvantaged by introducing employee-clinics, since employees would be treated more swiftly than non-employees. These opponents of employee-clinics purely focus on the deterioration of the relative access to health services rather than on actual access, since the worsening of the relative position is

accompanied by an improvement of the absolute position of non-employees. If employee-clinics would be allowed, some employees would disappear from these regular waiting lists and would be treated in the evening or in the weekends. Hence, regular waiting lists would shrink and non-employees would receive care more swiftly during regular hours than before.

8.4 Employee-clinics and efficiency

To evaluate the efficiency of employee-clinics, the traditional welfare economic criterion of Pareto may be applied. According to the Pareto-criterion any allocation of goods is efficient if no re-allocation can make somebody better off, without making somebody else worse off. If this point is reached the position is called Pareto-optimal. Changes in society can also be judged by using this criterion, since it is possible to look whether a change makes people better or worse off.

This may be illustrated by a simple example. Consider three possible allocations of goods in society (represented here as one number) as shown in Table 8.2. We will assume, for sake of simplicity, that interpersonal utility comparisons are possible and that utility derived from goods are equal to the number of goods received.

	Distribution I	Distribution II	Distribution III
Group A	10	18	12
Group B	5	2	6

Table 8.2: Example of allocations

All of the distributions in Table 8.2 may be Pareto optimal. The Pareto criterion does not normatively appraise the initial distribution of goods, i.e. the distribution of wealth at the point of consideration is taken as a fact and not questioned. So, even though distribution II in Table 8.2 may be considered inequitable, it still is Pareto-optimal if redistribution of goods from the rich to the poor would make those rich worse off. Sen (1970)

states: 'In short, a society or an economy can be Pareto-optimal and still be perfectly disgusting'. Therefore, when appraising changes, the starting point is crucial, but not subject to any normative valuation. Changes from one distribution to another may be evaluated by the Pareto criterion.

Moving from distribution I to II in Table 8.2, total wealth (A + B) increases, however, group B is worse off (2 instead of 5 units) which is not Pareto-optimal. Moving from II to III neither is Pareto-optimal as it makes group A worse off, although the inequality decreases. Moving from I to III can be considered Pareto-optimal, since both parties gain; no one is worse off.

The introduction of employee-clinics may be categorized as a Pareto-optimal change. This is the case because both the employees and non-employees will benefit from this introduction, since both groups will be treated sooner because of the expansion of capacity. The fact that employees will benefit more from such a movement than non-employees is not relevant. As for the change from distribution I to III in Table 8.2, it does not matter that group A gains 2 and group B gains 1, it only matters that both gain (or for Pareto optimality that no one looses). Employee-clinics may therefore be considered to be an efficient way to invest in health. This is especially true, since the rationale for employee-clinics is that the costs for priority care are less than the savings made on production losses and/or replacement costs incurred by firms. This rationale also invalidates the objection that firms would increase their production costs and therefore costs of priority care would finally be paid by employees and non-employees themselves.

Nord (1990) also investigated the benefits from giving priority care to employees on sick-leave. He found that it indeed could be possible to reduce waiting lists for all by giving priority to employees on sick-leave, by re-allocating the production gains from reduced absence from work by giving employees priority *within normal working hours of the hospital.* The solution of creating additional supply by working overtime or by 'buying services from another hospital' can also be 'highly justified economically' (Nord, 1990). An important issue, also raised by Nord is whether increased supply of health care will increase the demand for health care. If increased supply creates it's own demand some of the positive effects from the initial waiting time reduction may disappear because increased demand will

increase waiting time. Nord feels that a 'hospital should in principle make up its mind as to whom an increase in supply is supposed to benefit.' In other words, if the criteria for admission to a waiting list would remain unchanged, an increase in effective and rewarded demand should not have to occur. Hence, a uniform categorisation of diseases and severity and more comparative information about waiting lists is indispensable. If demand indeed increases, the extent to which waiting lists and waiting time return to their initial levels is determinant for whether the gains can cover the costs of increased supply.

8.5 Employee-clinics and equity

Whether it could be considered equitable to give priority care to employees is a much debated issue in the Netherlands. Many policymakers argue that giving priority for other than medical reasons could not be tolerated because it would create a two-tier system in health care. Some even consider employee-clinics to be in conflict with the Dutch Constitution in which the principle of nondiscrimination is included. In this section we will consider whether an introduction of employee-clinics can be justified on equity grounds and whether 'problems of justice and fairness ... might be solved by assuring that part of the production gains were used to increase treatment capacity to the benefit of non-sick-leave patients also' (Nord, 1990).

Rawls' theory of justice

There is no positive general rule on what is just and what is not. Hence, every justification of any change in society is a normative one. Even the interpretation of different theories of justice is not always straightforward. Here we will use an important view that was put forward by John Rawls in his Theory of Justice (1971). Rawls argues that social contracts made under the veil of ignorance may be considered just. Under the veil of ignorance people are unaware of their social status, capacities, place of birth, et cetera. Therefore, they will not act to enhance the wellbeing of one particular group, namely the one to which they belong, as normally observed in society, but they will enhance their wellbeing, not knowing to which group in society they will belong. According to Rawls, people under the veil of ignorance will judge different allocations of goods by the maximin-rule. This rule means that people under the veil of ignorance try to maximize the

position of the people who are worst off in society. They will maximize the wellbeing of people who have the minimum wellbeing. This risk-averse behavior occurs because people under the veil of ignorance may still end up in the group of people that are worst off in society.

Decision-making under the veil of ignorance may be illustrated by using the different distributions in Table 8.2. Using the maximin criterion of Rawls one would have to choose distribution III in this example. Then, the group with least assets, group B, at least receives the maximum of the three distributions possible, i.e. 6. Although under regulation I the difference between the two groups is smallest and although under regulation II total wealth is highest, the only criterion here is the position of group B, which needs to be maximized.

If such a construction would be used to judge the desirability of introducing employee-clinics, it seems obvious that if the non-employees can indeed benefit from this introduction by a reduction of waiting time during regular hours, then this would enhance their position. In terms of Table 8.2, this could be illustrated by a movement from distribution I to distribution III. The fact that employees (group A) benefit more from the introduction, since their waiting time is reduced more, is not relevant here. Therefore, using Rawls' concept of justice, the introduction of employee-clinics may be considered to be just.

One might argue that Paretian, Rawlsian and strict egalitarian approaches are rather extreme. In reality decision-makers might prefer trading-off equity and efficiency. One possible way of doing so is to require employers to pay an additional 'solidarity surcharge' directly to the hospital when using employee-clinics (Brouwer et al., 1996). The funds that are collected in this way, could be used to reduce regular waiting time. Such a cross-subsidy would improve the position of non-employees even further.

Rawls uses two principles in his theory of justice. The first principle states that everyone should have equal rights to the highest degree of basic freedom, while considering the same freedom for others. The second states that inequalities must be beneficial to the worst off in society and that there should be equality in opportunity (especially to reach higher public positions). The first and the second principle are to be satisfied in a serial

order, with the first principle prior to the second. If people have the same basic rights, as we believe is reasonably satisfied in most Western countries, then an increase in inequality can be justified when the worst off benefit from it. Some authors claim that health care services belong in the group of basic goods which should be distributed equally among all, or at least that differences should based solely on medical need (Daniels, 1985; Seeverens, 1997; Oosterling-Schiereck and Dillmann, 1997). However, placing access to health care services under the first principle without ever reaching the second does not seem an appropriate interpretation of Rawls' principles. This holds because it totally ignores his struggle to avoid pure egalitarian decision-making, which often denies people any improvement. If equality (the first principle) is applied that strictly, then the second principle becomes meaningless for all decisions about (re-) allocation of health care resources (Brouwer et al., 1997c). Folland (1990) criticizes the proposition of Daniels (1985) that health care should be distributed on the basis of medical need solely. Needs are not uniform and some 'objective' ranking of needs of different persons has to be found. Moreover, according to Folland, equal opportunity for all is not only achieved by providing health care, but by numerous other scarce goods in society. Finally, Folland points at an apparent inconsistency in Daniels' reasoning, where he admits that health care spending has to be weighed against funding other important social institutions. If such weighting should occur, priority care for employees may even be acceptable from Daniels' point of view, because it seems possible to simultaneously increase the level of social benefits for all dependent on social insurance *and* the available funds for health care.

Furthermore, Daniels' fair equality of opportunity provides another rationale for the introduction of employee-clinics because they may weaken incentives for employee selection in the labor market (Brouwer et al., 1998b). In the Netherlands, an investigation by the supervising body for social insurance provided evidence of firms starting to increasingly select healthy employees in order to reduce absence due to illness (CTSV, 1995b). Job applicants that had a medical record which led employers to believe that they might be absent more than average were turned down. The government responded to this development by prohibiting mandatory medical tests as part of a selection procedure. However, such a formal prohibition can easily be countered by healthy applicants voluntarily agreeing to medical tests or by implicit medical tests, like appearance and

age. A recent investigation confirms that firms are indeed increasingly engaged in selection of healthy workers (Willems, 1998). Such an increase in selection 'at the gate' may increase social inequity between healthy and unhealthy workers. A reduction of waiting time for health care may decrease the incentives for selection. Hence, while employee-clinics may cause a different treatment of employees and non-employees in the health care market, it may reduce the incentives for different treatment of more and less healthy persons in the labor market. In that sense reducing incentives to select healthy potential employees will enhance the opportunities of certain groups in society, an aspect that Daniels puts in the centre of his reasoning.

8.6 Relative versus absolute differences

An important objection to the use of the Pareto-criterion for efficiency could be that this criterion only considers the absolute level of care, and not the relative level. As pointed out by Rice (1997) one may question whether the use of absolute levels of wealth is a good way of representing utility that people derive from that level. Rice insists that relative levels are at least as important as absolute levels. This implies that neglecting the fact that employees benefit *more* from the introduction of priority care than non-employees, leads to wrong conclusions about the efficiency of such a development. Non-employees would not care as much about their absolute gain, but more about the absolute difference between themselves and employees. Pauly (1997) in a reply to Rice states:

'*... consider ... envy as affecting the evaluation of a costly new therapy that only the wealthy choose to buy. If we postulate envy as motivation of on the part of non-rich, it is true that this technology will make them worse off; still, it is preposterous to propose forbidding technologies that make one envious person just a tiny bit more jealous.*'

In real life, however, politicians probably cannot ignore social envy in their decision-making. In fact, the same reasoning on relative versus absolute changes may be considered for equity arguments as well. Many propose an egalitarian principle to be appropriate for the allocation of health care (e.g. Daniels, 1985). The only basis for difference in treatment ought to be medical criteria. Equal treatment for equal (medical) need is the underlying notion. This means that the Rawlsian focus on the absolute level of the

worst off in society is inappropriate. Rather, society's aim should be to minimize difference in treatment for people with the same medical need. However, the price of equality is equal here to denying both employees and non-employees an improvement in relation to the previous situation, i.e. a reduction of their waiting time. In fact, Rawls' theory of justice is a means to avoid such a strict egalitarian reasoning which does not allow for any improvement if it is not fully redistributed.

In the Dutch health care system arrangements in which firms play an important role and explicit priority is given based on non-medical criteria (by using capacity that remained unused before), are rather new. The egalitarian principle of equal treatment for equal need was applied (at least formally) quite strictly. But although explicit priority care is dismissed, implicit priority care is quite common. There is an abundance of anecdotes about people who jump the queue: politicians, top athletes, relatives of physicians or people that cannot be missed at home. Although such anecdotal evidence is widely known, it never raised substantial opposition among politicians and the public. One might claim that there may exist some societal acceptance for top-athletes to be treated with priority because they are admired and their work brings joy to many. However, in a similar argument, one might claim that there may also be acceptance for workers, that cannot be missed at work, to be treated with priority, because they contribute to the national income, which benefits other people through re-distributions of wealth. If social envy would indeed be important, it is difficult to envisage why priority for workers is so much worse than for other groups.

8.7 Long-term versus short-term considerations

Although employee-clinics may offer a good solution to the problem of waiting lists in the short run, a prolonged reliance on such clinics may not be without risks. It is not beyond imagining that the introduction of employee-clinics in the longer run may result in a further shift from public to private financing of health care delivery. In other words, in the long run, the government may cut back on the regular budget, i.e. public investment in health care, because of the additional private investments. This will lead to a reduction of the regular budget and therefore to a decrease in the

absolute level of access to health care services for people relying on the public sector, especially non-employees. Also, for employees, the level of care may become dependent on fluctuations in economic development. Although one might claim that a government that does not want to allow a relative deterioration of access to health care services may not act as described above, it should not be ignored that these risks are present. A crucial test of government behavior is that without curtailing public resources employee-clinics will be a self-limiting phenomenon, because employers would only be willing to invest in priority care as long as the reduction in waiting time is sufficient to cover the costs. So, if employee-clinics continue to grow this may be a sign of a reduction of public funds for which the government can be held responsible.

8.8 Conclusion

Employee-clinics are a possible way to attract additional funding for health care, specifically aimed at a group of waiting patients whose time is valuable for a third party (employers) that is willing and able to pay for expanding (the use of) hospital capacity out of savings on avoided production losses. In this way it is possible to improve access for both employees and non-employees but to a different degree. According to Pareto's welfare criterion of efficiency and Rawls' theory of justice such an improvement can be considered both efficient and equitable. Although these criteria are well-known and broadly accepted, one should recognize that other perspectives may yield different conclusions.

Of course, an ultimate test whether these considerations are normatively valid is the acceptance of employee-clinics by politicians and society, when fully informed about the (potential) absolute gains for all involved parties (Culyer and Evans, 1996). If even then politicians and society refuse to admit at least some pilots with employee-clinics, normative decision rules about *relative* positions of employees and non-employees apparently are more important than *absolute* ones. In other words, social envy may be that strong as to prevent an absolute improvement for all parties involved. Disguised by the veil of ignorance, however, priority care for employees, may well turn out to be a blessing.

9. A fair approach to discounting future effects: taking a societal perspective

Summary

It is often recommended to discount future costs and health effects at the same rate, usually 5%. In this chapter it is argued that from a societal perspective the current discount procedure may be inappropriate for health effects. Instead of implicitly basing the discount rate for health effects on that for costs, it should be based on the real change in the (societal) valuation of health effects over time. There may be a change in this valuation due to increasing life-expectancy over time and diminishing marginal utility of additional health effects.

9.1 Introduction

In economic evaluations discounting is an issue of considerable debate when health programs yield costs or health effects in the future. Normally comparisons between different health programs are made by way of cost-effectiveness ratios containing present values of future costs and effects. Correcting costs and effects for the time they occur by weighting them to calculate a present value is called discounting;. However, which discount rates one should use is not an easily answered question, nor whether a constant discount rate should be applied (e.g. Harvey, 1994; Cairns and Van der Pol, 1996). In recently published guidelines for cost-effectiveness analysis (Gold et al., 1996; Drummond et al., 1997) the most prominent view is that one discount rate should be used both for costs and for effects.

The necessity of calculating a present value of future costs is not much debated within the framework of CEA, although the exact rate to use is. Discounting future lives saved on the other hand may be considered

inappropriate because it may not be that clear why the value of lives saved in the future should be valued less than those now. This is especially the case when dealing with future generations. This chapter provides a theoretical background for addressing these issues from a societal perspective using an extra welfarist approach and (a relaxed version of) the Rawlsian theory of justice (Rawls, 1971).

The outline of this chapter is as follows. First, in section 9.2 the current views on discounting are discussed. Section 9.3 discusses how taking a societal perspective in CEA may be operationalized. Section 9.4 elaborates on how health effects may be discounted from a societal perspective, while section 9.5 indicates a preliminary solution for presenting future effects and future research areas.

9.2 Current views on discounting

The procedure of discounting, reflecting a lower weight attached to things that happen in the future than those happening now, can alter the results of an economic evaluation substantially and therefore also the subsequent decision-making process (Parsonage and Neuburger, 1992). When costs or effects occur far in the future the effect of discounting becomes apparent. To illustrate this: discounting future effects with a discount rate of 5% implies that a life year in 40 years counts as only 1/7th of a life year gained now in the economic evaluation (because $1/(1.05)^{40} = 0{,}149$ or about 1/7th). Thus, a vaccination program, with costs in this year and health gains in the future will be seen as less cost-effective than a program costing the same but yielding the same amount of health effects this year, because the future effects receive less weight.

There is, at least at the practical level, a broad consensus that one discount rate (generally 5%) should be used both for costs and effects (Gold et al., 1996; Drummond et al., 1997). This has been supported by three economic arguments all of which can be challenged.

First, it is clear that 'economic theory implies that in a perfectly competitive, risk-free, tax-free world in which all commodities (including something called 'health') are perfectly divisible - so that individual

decision-makers could precisely adapt their consumption of goods and services over time - there would be but one discount rate.' (Gold et al., 1996, page 216). In this situation therefore, the discount rate for costs would be the same as for effects. However, these conditions are clearly not met in reality, certainly not in health care.

Second, there is the so-called postponing paradox (Keeler and Cretin, 1983). Keeler and Cretin demonstrate that the use of different discount rates for costs and for effects (with the latter being smaller than the former), will lead to a situation in which infinite postponement of programs is always superior in terms of a cost-effectiveness ratio to performing program now. A postponed program will always have a better cost-effectiveness ratio than the same program performed now. Since no effects will be generated in this case, such an outcome can hardly be called optimal. A solution for the postponing paradox is the use of the same discount rate both for costs and for effects. Then, the postponed program will have an cost-effectiveness ratio equal to that of an identical program performed now.

It should be noted here that the postponing paradox is present in all consumptive activities - health care is no exception. Moreover, using the same discount rate for costs and effects changes a preference for postponing the program into indifference between postponing and performing a certain program now. This indifference is established at the cost of possibly misrepresenting (the value of) future effects. In addition, the postponing paradox is only relevant, when one has to decide whether to spend money now or later on a consumptive activity. When ranking programs within strict budgets that have to be allocated within one year, the paradox will not be relevant. These strict budgets are very common in 'real life' and therefore introducing different discount rates for costs and for effects is not expected to lead to infinite postponing in reality. Finally, the paradox does not necessarily imply that the discount rate for costs should also be used for effects. Using the discount rate for effects (whatever it may be) for costs as well will also solve the problem of the paradox and may be considered equally inappropriate.

Third, the US Panel on Cost-Effectiveness in Health and Medicine and also Drummond and colleagues present the 'consistency argument' of

Weinstein and Stason (1977) as an important foundation for the recommendation to use one discount rate for both costs and effects. Weinstein and Stason have demonstrated that inconsistencies may occur when discounting at two different rates for effects and costs. The crucial assumption underlying the reasoning of Weinstein and Stason is that 'life years are valued the same in relation to dollars in the present as in the future'. Thus a 'constant steady-state relation between dollars and health benefits' is assumed and 'opportunities for purchasing health benefits for dollars do not change over time'. This implies that there should be equal priority for programs that cost the same and yield the same benefits in a certain year regardless of what year. The only way to reach this is to use one discount rate for costs and for effects.

However, Van Hout (1998) has demonstrated that the assumptions of Weinstein and Stason are too restrictive and in fact equal assuming that there should be the same discount rate for costs and for effects. Also, note that Weinstein and Stason explicitly state that, although they propose the method as described above, they do not want to claim that, '... in any absolute sense, a year of life in the future is less valuable than a year of life in the present.' One might add that a future life year should not be considered less valuable in any relative sense either, without justification. Such a justification is not provided by Weinstein and Stason. Therefore, the need for one discount rate cannot be based on the consistency argument of Weinstein and Stason. As Broome (1994) puts it: 'It is perfectly consistent to discount commodities and not wellbeing.'

If we accept that the necessity for using one discount rate for both costs and effects does not exist, we face the difficult task of finding an appropriate rate for discounting costs and effects. Both should be discounted at their own appropriate rate, based on expected growth rates, that is, how fast income and (quality adjusted) life expectancy are expected to increase over time, and consequently on the marginal valuation of additional costs and effects. For effects this means 'combining the expected growth rate with an estimate of the marginal utility of the additional life years' (Van Hout, 1998). This outcome implies that the weight of future effects depends on the relative valuation that a future group of persons or society will have for these effects. Although

Van Hout does not explicitly define the societal perspective, his proposal may be considered to be more in line with taking a societal perspective.

9.3 Taking a societal perspective

The societal perspective is normally considered the proper perspective from which to conduct a cost-effectiveness analysis. By setting one standard perspective, the comparability of different cost-effectiveness analyses is increased. Gold et al. (1996) assert that 'when a CEA is conducted from the societal perspective, the analyst considers everyone affected by the intervention and counts all significant health outcomes and costs that flow from it, regardless of who experiences the outcomes or costs'. This will ensure that no significant aspects of the decision are left out of the analysis. In addition, the fair representation of all costs and effects is facilitated by choosing the societal perspective. 'One way to see the desirability of the societal perspective - of giving fair weight to all individuals and to all activities - is to imagine that we are looking at the world before we are born... and to ask what kind of world we would like it to be.' (Gold et al., 1996, page 6) This construction, of 'looking at the world before we are born' is much like the veil of ignorance as used by Rawls in his 'Theory of justice' (1971). This is a hypothetical situation in which persons have to decide on a social contract for the world they will occupy. These persons are unaware of the social status, capacity and health state they will have and unaware of the place and time at which they will be born. According to Rawls, contracting under this veil of ignorance can be considered fair. In this state the decision-makers do not give 'any weight to mere position in time.' (1973, page 294) This is because 'the persons in the original position have no information as to which generation they belong' and 'these broader restrictions on knowledge are appropriate ... because questions of social justice arise within generations as well as between them' (1973, page 137). Thus fair representation of the interests of all persons affected by the implementation of a program seems to be part of taking the societal perspective and this is expressed by not giving any weight to mere position in time. Note that this means that at first our aim should be to represent the effects yielded from a program in a fair way.

Gold et al. (1996) feel that current discount practice of using the same discount rate for costs and for effects can be considered fair, because it leads 'to resource allocations in a 'time neutral' fashion'. Time neutrality is assumed to exist because 'potential program beneficiaries who are identical in every respect except for their positions in time relative to the moment the decision-maker must act will receive equal treatment' (page 221). This holds because nominally equal programs receive equal priority, since:

$$(1)\ \frac{C_0}{E_0} = \frac{\frac{C_{40}}{(1+i)^{40}}}{\frac{E_{40}}{(1+r)^{40}}} = \frac{C_{40}}{E_{40}} \quad \text{if } i = r$$

in which C_0 and E_0 are the costs and effects in this year, C_{40} and E_{40} are the nominally equivalent costs and effects in year 40 and i and r are respectively the discount rate for costs and effects. However, nominal comparison of programs may lead to a serious misrepresentation of the impact of these programs. Rather, we should look for changes in the societal valuation of health effects and costs and give equal priority to programs that are equal in marginal societal valuation of costs and effects.

If current and future beneficiaries of programs are indeed identical, there is no reason to assume that they will not value a health effect identically. However, this fact does not necessarily have to be accompanied by the fact that society derives equal utility from a nominally equal amount of money now or in 40 years. As Van Hout (1998) indicates, this depends on the respective growth rates of (quality adjusted) life-expectancy and national income. Since the societal valuation of a life year saved may be the same for both persons or groups, while that for costs may be not due to rising national income over time, giving nominally equal programs at different points in time equal priority seems inappropriate as a general rule. Fair representation of effects may rather be interpreted here as representing the true societal valuation of effects.

Within an 'extra-welfarist' framework, the goal of CEA may be perceived to be the maximization of health benefits with the available budget. Therefore, society wishes to maximize the quantity of effects, for instance QALYs, though maybe subject to certain equity goals. In order to ensure that the operationalization of CEA is still in compliance with fair representation and maximizing health benefits, we have to ensure that the method for aggregating effects over time reflects the true change in the societal valuation of effects over time.

Taking the societal perspective also means abandoning purely individualistic approaches to CEA in general and discounting effects in particular. As indicated above, taking a societal perspective does not only imply incorporating the interests of all people now into the analysis in a fair way, but also that of future generations. If we were only dealing with individuals living here and now, we might consider examining individual preferences on the timing of health effects. For instance, the US Panel states that

'*if individuals place the same weight on future events as on those that will occur soon ... then a zero rate of time discount may be consistent with utility maximalization*' (Gold et al., 1996)

However, individual preferences regarding timing of effects are not the only relevant considerations. Measurement of individual discount rates for future health effects (e.g. Olsen, 1993; Cairns, 1994; Johannesson and Johansson, 1996) is not believed to be useful to serve as proxy for a societal discount rate, since 'the mere position in time' of the respondent is inevitably given weight in such measurement. Pigou (1932) has already argued that

'*there is wide agreement that the State should protect the interests of the future in some degree against the effects of our irrational discounting and of our preference for ourselves over our descendants*'.

Even within one generation one might claim that the time preference of individuals should be overruled by a societal time preference. A societal decision-maker may decide that 'statistically certain' effects in the distant future are just as important as those generated in the near future.

Although individuals may be influenced by pure myopia and fear of death, a social decision-maker may be assumed make more rational decisions.

An attempt to combine individual time preference with a societal rate of time preference has recently been discussed by Gyrd-Hansen and Søgaard (1998). This two-stage model for discounting was earlier introduced by Lipscomb (1989) and Gold et al. (1996) also discuss this method of discounting. Two discount rates are used in a two-stage method: an individual rate based on time preference with which a person calculates the present value of a stream of health effects and a societal discount rate for shifts in this 'present value of health effects' over time. Consider the following example: two persons of 40 receive a treatment which will give them an effect in three years. The first person lives in 1998 and will receive the effect in 2001, while the other person lives in 2020, receiving the effect in 2023. Time neutrality could be interpreted as meaning that we now should calculate the present value (in 1998 and 2020) similarly, for instance using individual time preferences (as it are indeed two individuals receiving the effects). A societal time preference may be used for discounting the present value (in principle equal in 1998 and 2020) from 2020 back to 1998. This societal discount rate should then be based on real differences expected over time between people. The two-stage discount approach may be linked to the work of Daniels (1988), who argues that people should be treated the same at similar points of their life, rather than in time. A 70 year-old person now should receive the same treatment as a 70 year-old person in 20 years, all other things equal.

In assessing the two-stage approach Gold et al. (1996) conclude that

'*the two-stage approach has drawbacks, not the least of which is that it depends on a basically arbitrary definition of the starting date of a program and it assigns different weights to health consequences occurring in different cohorts at the same point in calendar time*' (page 230).

In addition, individual time preference which is still used to some extent and will influence the outcomes of an analysis, can be considered irreconcilable with truly taking a societal perspective. Indeed, on a societal level one may wish to be less influenced by individualistic myopia and fear of death, since on a societal level a decision-maker may know with greater

certainty that some group of individuals will indeed receive these benefits. Why should a societal decision be influenced by the fear of some individuals of not receiving a benefit, when we know with certainty that some individuals will receive them? Rising above some individuals' uncertainty as to whether or not they will personally benefit 'is naturally society's viewpoint, since statistical life years will be consumed with certainty by some one in the population' (Gyrd-Hansen and Søgaard, 1998). The question therefore becomes, how to discount from a more strictly applied societal perspective.

9.4 Societal discounting?

A stricter way of discounting health effects from a societal perspective would imply that future generations should be protected from 'the effects of our irrational discounting and of our preference for ourselves over our descendants' (Pigou, 1932) and that current generations should be protected against their own 'irrational discounting' as well. Individual attitudes to risk and fear of death may be considered irrelevant in societal decision-making. At a societal level, future effects will be generated in some individuals and therefore decisions can be based on a more rational calculation. The societal decision-maker thus overrules the individual 'time horizon' and bases discounting solely on real differences between people over time that change the relative valuation of health effects. Such an approach is quite similar to the reason for discounting discussed by Van Hout (1998). Since decisions on the acceptability of a new drug or program are made at a societal level, it seems appropriate to use this strict societal perspective in the methods of CEA. Within this approach, future effects (and costs) have to be incorporated in a fair and realistic way. This means that the absolute amount of future effects should only be adjusted for their timing when obvious differences between people at different points in time are present or can be rightfully expected. It may be considered fair to adjust for these changes over time in so far as this adjustment reflects a difference in the valuation of these effects by those receiving them. The current method of discounting is, however, not based on differences in the valuation of effects over time.

Operationalization of such a discount procedure first depends on the possibility for finding an objective measure for relevant differences between people over time. One obvious difference that may serve as a basis for discounting future effects is a priori life expectancy (i.e. without implementing the program investigated). Changes over time in a priori life-expectancy can be used in an efficiency and an equity argument to discount future effects. In terms of efficiency, diminishing marginal utility from additional life years may be assumed to exist. Thus, adding a QALY to someone with a higher a priori life expectancy yields less utility than adding a QALY to someone with a low life expectancy.[17] The same argument can be used for discussions on distribution within one year or generation (Brouwer and Van Hout, 1998). If a change in life expectancy between people over time may be rightfully expected, a difference in the valuation of effects over time may be appropriate. This holds because a person living now, who has less to begin with, will benefit more from the effect relative to a person in 40 years time, who will have a higher life expectancy to begin with. This changing valuation of additional health benefits over time may then be seen as a basis for discounting health effects over time. This option appears quite in line with the suggestions made by Van Hout (1998) who argues that a difference in marginal (societal) utility from additional QALYs, due to increased possibilities for generating health benefits and, through that, increasing a priori life expectancy, is a reason for discounting future effects. The reason for discounting is thus related to real changes over time.

One might also consider it to be more equitable to attach more weight to a QALY gained in someone with a low a priori life-expectancy than in someone with a high a priori life-expectancy. Such a way of looking at differences over time seems in line with taking a societal perspective. Indeed, Rawls argued that under the veil of ignorance people would decide according to the maximin rule (1971), implying that the position of the worst-off in society would be maximized. The distribution of health services may, therefore, be aimed at maximizing the position of the worst-off. However, this strict rule implies that if future generations are expected to be better off than the present one, all health care spending should be aimed at the present generation, at least until the point is

[17] Note that the same argument may be used regarding the distribution of effects within one year or generation, e.g. Brouwer en Van Hout (1998)

reached in which both generations have, for instance, an equal (quality adjusted) life expectancy. Less strictly applied, one might interpret the maximin-rule as attaching *more* weight to an improvement of the worst-off in society or in time in stead of *all* weight. Indeed, one might claim that an equity-efficiency trade-off exists with inequalities in health over time as well as between people. This leads us to the interesting option of 'mildly applying' the maximin-rule as a basis for discounting in economic evaluation, similar to the fair innings argument (Williams, 1997). Williams asserts that resources be allocated according to the principle that people who have not yet reached their fair innings (younger persons) receive more weight in resource allocations than those who have. A similar technique could be applied to distributions over time.

9.5 Conclusions

The current discount procedure in economic evaluation can influence results considerably, while it is not necessarily the most appropriate procedure from a societal perspective. Taking a societal perspective, it seems more appropriate to discount future effects solely on the basis of real differences between people over time. Increasing life expectancy and therefore decreasing marginal valuation of additional QALYs over time may serve well as a basis for discounting future effects from a societal perspective.

Since at this point it is unclear exactly how the valuation of additional QALYs changes over time, it seems sensible for analysts to avoid discounting effects. Health effects should be reported undiscounted, mentioning when they will occur (Drummond et al., 1997). For calculating cost-effectiveness ratios it seems sensible to also use undiscounted health effects in base case analyses, demonstrating the effect of different discount rates for effects in a sensitivity analysis. Ultimately, policy makers will have to decide on how to weight the timing of effects in CEA, considering fairness, efficiency and the interests of present and future generations. The recommendation of Krahn and Gafni (1993) that the choice for one specific discount rate should be the result of a 'political' consensus, in which all relevant aspects may be considered seems appropriate. Future research should be aimed at finding ways to

transform changes in the valuation of health effects due to changes in life-expectancy over time into a discount rate for effects. That information may be used to better inform policymakers on the general devaluation of additional health effects over time.

10. Discussion

Summary
This chapter presents a discussion of the main conclusions that can be drawn from the previous chapters with respect to the central question of the thesis: How does taking a societal perspective in economic evaluations of health care, following a decision-maker's approach, change methodology, especially those methods concerning time and time costs. Specifically, it will draw conclusions on the subjects of patient time in economic evaluation, informal care and discounting. Furthermore, some areas for future research are pointed out.

10.1 The societal perspective and economic evaluation

Mostly, it is recommended to conduct an economic evaluation from a societal perspective, incorporating all costs and health effects, regardless of who bears these costs or who experiences these effects. This should ensure that all significant changes both in terms of costs and health effects will be part of the analysis and the final policy recommendations. To take such a starting point in economic evaluation has consequences not only for what to measure in the evaluation, but also for how to measure certain quantities in an evaluation.

When taking a societal perspective it is important to consider that certain gains and losses may level out on a societal level. For instance, when considering productivity losses, it is not enough to look at the production loss of the patient, but one also has to consider the production gain of a person possibly replacing him or her. That way, all impacts on a societal level may be incorporated and balanced, thus ensuring that in an evaluation from a societal perspective one 'never counts as a gain what is really someone else's loss' (Gold et al., 1996, page 7).

A related debate is that on the general approach to economic evaluations of health care programs in general and cost-effectiveness analysis in particular, i.e. the choice between a welfarist and extra-welfarist view. Welfarists follow a more strict welfare theoretical approach to (the methodology of) economic evaluations, while extra-welfarist take a more broad and perhaps pragmatic view on economic evaluation, what we have called in chapter 2 the 'decision-maker's approach' (DMA). Although both welfarists and extra-welfarists aim at aiding policymakers in decision-making on health care services, the differences in approach cause different methods and methodological choices to be advocated. Welfarists often use optimalisation of more individualistic micro-economic models as a basis for methodological recommendations, e.g. the recent work by Garber and Phelps (1997) and for instance the human capital method for estimating productivity costs. However, arguably a gap exists between these simple individualistic models and real life decision-making on a societal level, which makes it difficult to use these models and their implications in real life. The strict focus on utility in welfarism may also be perceived to be less useful for decisions in health care, where the aim may be more to produce health or restore capabilities to live as one would like to. In that sense, the aim of the DMA, the approach chosen in this thesis, may be seen as aiding policymakers in optimizing (weighted) health benefits from a given budget.

Furthermore, a fair representation of costs and effects in principle is part of the societal perspective, as discussed in chapter 9. 'One way to see the desirability of the societal perspective - of giving fair weight to all individuals and to all activities - is to imagine that we are looking at the world before we are born... and to ask what kind of world we would like it to be.' (Gold et al., 1996, page 6) This 'veil of ignorance'-approach, used in chapters 8 and 9, is a hypothetical construct in which persons have to decide on a social contract for the world they will occupy, unaware of their final social status, capacities, health state, time of birth, et cetera. According to Rawls contracting under this veil of ignorance can be considered fair (1971). This veil of ignorance approach has relevance for questions of fairness within economic evaluations, both for specific studies (e.g. a study into the consequences of priority care for employees)

and in methodological matters (e.g. in deciding on an appropriate discount rate for future health effects).

It is especially important to note that the choice for taking a societal perspective on a general level, needs translation to the practical methodological choices in an evaluation. In order to be able to ensure that results can be interpreted from a societal perspective, methodologies used in an analysis have to adhere to this perspective as well.

10.2 Patient time

The valuation of changes in time use of patients is an area where taking a societal perspective has clear consequences for the methodology used in economic evaluations. The traditionally used human capital method may be considered too individualistic in perspective to serve well as a tool for valuing production losses from a societal perspective. Indeed, normally encountered replacement mechanisms need to be taken into account when assessing production losses in case of absence from paid work and disability. The friction cost method is at this moment the only practical method that does so. In that respect, the friction cost method may be seen as an inherently societal valuation method for production losses.

An important aspect of patient time is the complementarity of quality of life measurement and cost-components. Not only may one prefer to measure certain aspects of time use in terms of quality of life, like changes in leisure time, but one also has to consider to what extent these aspects are already captured in terms of quality of life in frequently used QALY-measures like the EuroQol/EQ-5D instrument. To neglect the QALY-implications of changes in time use may lead to double-counting, i.e. incorporating a quantity both in terms of health effects and in terms of costs in an economic evaluation, or to neglecting certain quantities. It seems recommendable to incorporate the impact from impaired role functioning both for paid and unpaid work and changes in the ability to engage in leisure in terms of quality of life, as elaborated on in chapter 3. The output losses for paid and unpaid work may be captured in monetary terms. It is stressed here that the often encountered emphasis on production losses related to the impaired ability to perform paid labor

should not be interpreted as if unpaid labor is unimportant. Much effort is needed to further investigate illness in unpaid work and related replacement issues. This may be performed by interviewing patients who's primary time use is unpaid work, focusing on the replacement issues within the household. The results may lead to better valuation methods of unpaid work. Also, future research could be aimed at better examining the complementarity of QALY-instruments and costing principles, e.g. the (possible) tension between QALY-assessment by the general public and more patient specific cost-measurement. One way of doing so, is to examine the valuation of specific QALY-dimensions by different subgroups in society. These subgroups may be defined by using primary time-use (paid work, unpaid work or leisure) as an important classification-criterion. Another characteristic that may prove to be of importance is whether or not someone is involved in raising children. If differences in the valuation of certain dimensions of quality of life are found then more specific health state valuations may be used in economic evaluations of interventions that affect particular groups in society. A final point for further empirical investigation is replacement in paid work. The processes of making up for lost work by the patient or taking over by colleagues and the related costs need to be examined further.

Although there indeed seems to exist a relationship between changes in time use and quality of life, measuring and valuing production losses in terms of quality of life as proposed by Gold et al. (1996) is discouraged here for several reasons, as explained in chapters 4 and 5. One important reason is the weak relationship between income and productivity in many developed countries. Social and private insurance will mostly sustain a certain level of income even in case of full disability, which makes the relationship between productivity and income (and then quality of life) rather inappropriate for precise measurement. Also, the societal perspective is abandoned in such an approach as the personal income change is considered for as long as it affects the patient, while not considering the improvement for the person replacing him or her. Furthermore, one may question the appropriateness of incorporating productivity costs in the denominator of a C/E-ratio. In principle, the numerator should capture true societal costs while the denominator should be exclusively reserved for changes in (health related) quality of life. Further research on whether or not individuals take into account

income when assessing quality of life for different health states and how this affects the final results of an intervention is encouraged. This could be performed in a two-stage approach, first asking individuals to assess the quality of life of different health states and subsequently asking them whether or not they took into account income while assessing quality of life. In that way, it may be determined what differences occur when people do or do not take into account income when assigning quality of life-weights to different health states.

Measuring impairment of performing productive activities while being at work requires instruments capable to measure the level of performance during days ill at work accurately. As indicated in chapter 6, the instruments currently used provide different estimates of costs related to illness without absence. Additional comparisons of the results of different measurement methods, preferably to a gold standard (real production data on an individual level) are needed to further determine the validity of the different methods. Still, illness at work proved to be quite common with on average 7% of respondents (a group of employees belonging to a Dutch trade-firm) indicating on a workday that they felt ill while working. Since the level of functioning compared with normal is often reasonably high, the costs related to this common phenomenon are, on the level of a firm, consequently quite low. However, for specific patients groups these costs may be considerable. An interesting, yet virtually unexplored aspect of performance level is the difference between quality and quantity of output on days working with impairment. This should be further examined in the future, along with the different methods in general, while also better focusing on replacement by colleagues and the specific setting of work.

Production losses related to impaired functioning or absence from work may constitute a considerable societal loss. Each day that the illness continues to exist additional production losses may occur (at least within the friction period). Since employers have become responsible for the payment of sickness benefits in The Netherlands (and therefore are directly faced with increased costs or uncompensated production losses) they are more aware of this fact. One of the consequences of this awareness is the discussion about priority care for employees. As shown in chapter 8, by organizing 'employee-clinics' in such a way that they

increase (the use of existing) treatment capacity, priority care may be considered both efficient and equitable. Whether or not the benefits for both employees and others can be realized when introducing priority care for employees may be assessed by monitoring the waiting time of both groups after introducing priority care. Since over time the introduction of priority care will not be the only change in the health care system, it will be more difficult to determine the specific effects of introducing priority care. Also, when only few hospitals provide priority care, additional patients may visit these hospitals hoping to benefit from reductions in waiting time during normal hours.

Receiving priority treatment abroad in other EU-countries may not be that easy to prohibit, especially when considering the recent rulings of the European court (Hermans and Brouwer; 1998a; Hermans and Brouwer, 1998b; Brouwer and Hermans, 1999). Another point that is clear from the discussion of employee-clinics is that it is the use of a societal perspective that indicates misallocations between health care and other sectors. Through rationing of health care, resulting in waiting lists, money is saved in the health care system, but a larger amount of costs may be incurred in terms of production losses or increased social benefits. Using the societal perspective, such relationships may be pointed out and a better allocation of resources may be decided upon, which may benefit all parties. Future research may be aimed at assessing the costs and effects of quicker care (for employees and other members of society). A comparison of costs and effects related to a certain intervention, when performed after a normal waiting period and performed virtually without waiting, could provide valuable information.

An interesting discrepancy that may be detected here, is that although for economic evaluations the incorporation of indirect non-medical costs or productivity costs is now commonly recommended in methodological guidelines, a direct consideration of these costs in relation to the organisation and prioritisation of health care services, as performed in chapter 8, at least in the Netherlands leads to political and social resistance. So, although on a health program level we do allow indirect costs to play (a sometimes important) role in judging which programs should receive priority, a more direct 'favouring' of groups that are productive in society is considered unwanted. Future research may be

aimed at developing instruments capable of measuring the distributional effects of incorporating productivity costs.

10.3 Informal care

Informal caregivers give up on normal activities because of the incidence of illness in their direct social environment. Although difficult, the valuation of time of informal caregivers is important when evaluating for example new drugs for patients with diseases that are typically associated with a considerable amount of informal care, like Alzheimer's Disease (e.g. Busschbach et al., 1998). Therefore, better methods for valuing informal care are called for to ensure a good estimation of the true societal impact of informal care. The incorporation of informal care in economic evaluations of health care programs is generally performed by attaching some monetary value to the hours of informal care received. In chapter 7, some alternative monetary valuation methods for informal care were discussed: the market price method, the opportunity cost method or contingent valuation. Although, for instance the market price method is easy to use, and all three methods yield purely monetary results that are in principle easy to incorporate in the analysis, one may question whether they are the most appropriate to use from a societal perspective, because especially in long spells of providing informal care the consequences include effects like fatigue, giving up leisure activities and less social contacts, which are difficult (or impossible) to fully capture in monetary terms.

Since the societal perspective dictates incorporating all significant changes resulting from an intervention, regardless of who experiences these changes, the impact of interventions on the time use and quality of life or general well being of informal caregivers should not be neglected when these are (expected to be) significant. This is even more important considering the fact that a fair representation of all persons affected by an intervention is also part of taking a societal perspective. Measuring the impact in terms of 'care-related quality of life' as proposed in chapter 7, may provide an alternative way of measuring effects on quality of life, complementary to costing the time involved in informal caregiving. Similar to the way the time of patients is proposed to be valued, a

combination of quality of life and costing principles can be used for valuing the time of informal caregivers.

The measurement of informal caregiver quality of life does raise questions, not only about what to measure, i.e. pure effects from caregiving or more general 'family-effects', but also about how to incorporate the results in a cost-effectiveness analysis. The US Panel on Cost-Effectiveness Analysis has advocated 'to think broadly about the people affected by the intervention and begin to include health related quality of life effects of significant others in sensitivity analyses when they are important'. However, as long as general 'family-effects' are not measured in other evaluations as well, it is argued to focus on informal care-related changes in quality of life. The results may be presented separately in the analysis. Future research may be aimed at further developing and testing caregiver quality of life instruments that may be used together with the costing-principles of their time input. In addition, it is interesting to investigate how policymakers weigh informal caregiver quality of life in their decision-making. This may be further tested by examining trade-offs between patient and informal caregiver quality of life by these policymakers.

10.4 Discounting

At present, the most commonly used and promoted discount technique is to discount both costs and health effects at a 3 to 5% discount rate. However, as was argued in chapter 9, taking a societal perspective may be perceived as having consequences here as well. A fair representation of current and future health effects may be considered essential in taking a societal perspective. The current discount procedure implies a relatively strong preference for current effects and current generations over future effects and generations, by attaching only a small weight to longer-term future effects. Although this may be understandable from an individualistic viewpoint, it may be questioned whether societal decision-making should not abandon the present individualistic perspective and have a less biased view on the value of future effects. Following the justice-principles developed by Rawls, one may argue that a more 'timeless' or intergenerational societal perspective can be advocated, in

which the reasons for discounting effects are more objective, i.e. based on actual differences of persons over time. Increasing life-expectancy over time is such an objective difference and this, together with the principle of diminishing marginal utility of additional QALYs, may provide a better rationale for discounting health effects.

The current viewpoint, i.e. that it is necessary to discount both costs and effects with the same discount rate, is mainly based on three arguments, none of which seems enough justification for choosing a discount procedure implying a rapid devaluation of effects over time. From a societal perspective, it seems more appropriate to discount future effects on the basis of real differences between people over time. As further explained in chapter 9 increasing life expectancy and therefore decreasing marginal valuation of additional QALYs over time may serve well as a (theoretical) basis for discounting future effects. Although this is partly possible in the two-stage discount-framework, a more strictly applied societal approach may be preferred. It remains difficult however to give explicit guidance on which discount rate to use from a societal perspective, since it is unclear how the valuation of additional QALYs changes over time. Therefore, it seems recommendable for analysts to avoid discounting effects, reporting effects undiscounted, while mentioning when these effects will occur. Then policymakers will have to decide on how to weigh timing of effects in CEA, considering fairness, efficiency and the interests of present and future generations. Future research may be aimed at finding ways to transform changes in the valuation of health effects due to changes in life-expectancy over time into a discount rate for effects. That information may be used to better inform policymakers on the general devaluation of additional health effects over time. One way of further determining such a discount rate for health effects would be to determine which weight policymakers attach to QALYs gained in persons with different age-profiles. From those judgements a discount rate could be formulated (based on increasing quality adjusted life-expectancy over time), with which the policymakers could be subsequently confronted, in order to test whether they agree with the specified discount rate. Another option would be to test whether policymakers and the general public endorse the principle of discounting on the basis of a change in quality adjusted life-expectancy.

10.5 Epilogue

Taking a societal perspective in an economic evaluation of health care programs has consequences for its methodology. In this thesis some consequences concerning the valuation of patient time, informal care and discounting future effects in an 'extra welfarist' societal economic evaluation have been presented, three topics that are much debated in the literature, lacking consensus. The current differences in recommendations on these methodological issues may be traced back to a difference in 'tradition', i.e. whether a more strict welfarist approach is adhered to or a more broad extra-welfarist approach, and also to whether or not a societal perspective is consequently pursued, in all relevant methodological choices. In this thesis I have argued that certain methods may be less appropriate to use in this respect and have presented alternative ways of handling patient time, informal caregiver time and future effects in economic evaluations of health care.

Further harmonization of methodologies used in economic evaluations and the development of a uniform presentation of results are two of the challenges that lie ahead. Harmonization of methodologies will further enhance the comparability of different studies and the transparency of how results are derived. A uniform presentation of results can also improve the information-transferal from analysts to policymakers. It goes without saying that much effort is needed to further develop a complete 'toolkit' for societal economic evaluations and to further refine and test this toolkit. In that sense, I hope this thesis contributes to the development of and consensus on methods that inherently use a societal perspective.

References

Baldwin S and K Gerard, 1990, Caring at home for children with mental handicaps, in: Baldwin S, Godfrey C and Propper C (eds), Quality of life: perspectives and policy (Routledge, London)

Bleichrodt H, 1995, QALYs and HYEs: under what conditions are they equivalent?, Journal of Health Economics 14, 17-37

Bleichrodt H, 1997, Health utility indices and equity considerations, Journal of Health Economics 16, 65-91

Boyle MH, GW Torrance, JC Sinclair and SP Horwood, 1983, Economic evaluation of neonatal intensive care of very-low-birth-weight infants, New England Journal of Medicine 308, 1330-1337

Broome J, 1993, Qalys, Journal of Public Economics 50, 149-167

Broome J, 1994, Discounting the future, Philosophy and Public Affairs 23, 128-156

Brouwer WBF, FT Schut and FFH Rutten, 1996, Employeeclinics in light of Rawls and Pareto (in Dutch), Economisch Statistische Berichten 81, 973 –975

Brouwer WBF, MA Koopmanschap and FFH Rutten, 1997a, Productivity costs measurement through quality of life? A response to the recommendations of the U.S. Panel, Health Economics 6, 253-259

Brouwer WBF, MA Koopmanschap and FFH Rutten, 1997b, Productivity costs in cost-effectiveness analysis: numerator or denominator: a further discussion, Health Economics 6, 511-514

Brouwer WBF, FT Schut and FFH Rutten, 1997c, Reply (in Dutch), Economisch Statistische Berichten 82, 153-154

Brouwer WBF, MA Koopmanschap and FFH Rutten, 1998a, Patient and informal caregiver time in cost-effectiveness analysis. A response to recommendations of the U.S. Panel, International Journal of Technology Assessment in Health Care 14, 505-513

Brouwer WBF, FT Schut and FFH Rutten, 1998b, Waitinglists and employeeclinics (in Dutch), Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde 6, 51-54

Brouwer WBF and BA van Hout, 1998, Letter to the Editor: Diminishing marginal Utility and discounting future effects have similar consequences, British Medical Journal 317, 1155

Brouwer WBF and MA Koopmanschap, 1998, How to calculate indirect costs in economic evaluations, PharmacoEconomics 13, 563-566

Brouwer WBF and FT Schut, 1999, Priority care for employees: a blessing in disguise? Health Economics 8: 65-73

Brouwer WBF, NJA van Exel, MA Koopmanschap and FFH Rutten, 1999, The valuation of informal care in economic appraisal: a consideration of individual choice and societal costs of time. International Journal of Technology Assessment in Health Care 15, 147-160

Brouwer WBF and HEGM Hermans, 1999, Private clinics for employees as a Dutch solution for waiting lists: economic and legal arguments, Health Policy 47, 1-17

Brouwer WBF, MA Koopmanschap and FFH Rutten, 1999, Productivity losses without absence; measurement validation and empirical evidence, Health Policy 48, 13-27

Busschbach JJV, Brouwer WBF, Van der Donk A, Passchier J and FFH Rutten, 1998, An outline for a cost-effectiveness analysis of a drug for patients with Alzheimer's Disease, PharmacoEconomics 13, 21-34.

Cairns JA, 1994, Valuing future benefits, Health Economics 3, 221-230

Cairns JA and MM van der Pol, 1996, Saving future lives. A comparison of three discount models, Health Economics 6, 341-350

Canadian Coordinating Office for Health Technology Assessment, 1997, Guidelines for economic evaluation of pharmaceuticals: Canada. Ottawa, Ontario; 2nd edition

Centraal Bureau voor de Statistiek (CBS), 1997, Absence, results from a survey 1992-1995 (in Dutch), Sociaal Economische Maandstatistiek, April

Commonwealth of Australia, 1990, Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee: including submissions involving economic analyses. Department of Health, Housing and Community Services. Woden

CTSV, 1995a, Quantitative effects of financial incentives in the *Ziektewet* (in Dutch). Zoetermeer: College van Toezicht op de Sociale Verzekeringen

CTSV, 1995b, Risk-selection on the Dutch labourmarket (in Dutch). Zoetermeer: College van Toezicht op de Sociale Verzekeringen, 1995b.

Culyer AJ, 1990, Commodities, characteristics of commodities, characteristics of people, utilities and the quality of life, in: S Baldwin, C Godfrey and C Propper (eds), The quality of life: perspectives and policies. (London, Routledge)

Culyer AJ, 1991, The normative economics of health care finance and provision, in: A. McGuire, P Fenn and K Mayhew (eds), Providing Health Care: The economics of Alternative systems of finance and delivery (Oxford University Press, Oxford)

Culyer AJ and RG Evans, 1996, Mark Pauly on welfare economics: Normative rabbits from positive hats, Journal of Health Economics 15, 243-251

Daniels N, 1985, Just Health Care (Cambridge University Press, Cambridge)

Daniels N, 1988, Am I my parents keeper? An essay on justice between the young and the old (Oxford: Oxford University Press)

Daniels N, 1993, Rationing fairly: Programmatic considerations, Bioethics 7, 223-233

Dolan P, 1998, The measurement of individual utility and social welfare, Journal of Health Economics 17, 39-52

Dolan P, 1999, Drawing a veil over the measurement of social welfare – a reply to Johannesson, Journal of Health Economics 18, 387-390

Donaldson C, 1998, The (near) equivalence of cost-effectiveness and cost-benefit analysis: fact or fallacy?, PharmacoEconomics 13, 389-396

Doorslaer EKA van, A Wagstaff and FFH Rutten (eds), 1993, Equity in the finance and delivery of health care: an international perspective (Oxford University Press, Oxford)

Doorslaer EKA van, A Wagstaff, H Bleichrodt, S Calonge, UG Gerdtham, M Gerfin, J Geurts, L Gross and U Haekkinen, 1997, Income related inequalities in health: some international comparisons, Journal of Health Economics 16, 93-112

Drummond MF, 1991, Australian guidelines for cost-effectiveness studies of pharmaceuticals: The thin end of the boomerang? Centre for Health Economics Discussion paper 88 (University of York, York)

Drummond MF, EA Mohide, M Tew, DL Streiner, DM Pringle and JR Gilbert, 1991, Economic evaluation of a support program for caregiver of demented elderly, International Journal of Technology Assessment in Health Care 7, 209-219

Drummond MF, GL Stoddard and GW Torrance, 1987, Methods for the Economic Evaluation of Health Care Programmes (Oxford University Press, Oxford)

Drummond MF, B O'Brien, GL Stoddard and GW Torrance, 1997, Methods for the economic evaluation of health care programmes (Second Edition, Oxford University Press, Oxford)

Eddy DM, 1991, The individual versus society: resolving the conflict, Journal of the American Medical Association 265, 2399-2406

EuroQol Group, 1990, EuroQol: a new facility for the measurement of health-related quality of life, Health Policy 16, 199-208

Fitzgerald JM and A Gafni, 1990, A cost-effectiveness analysis of the routine use of isoniazid prophylaxis in patients with a positive mantoux skin test, American Review of Respiratory Disease 142: 848-853

Folland ST, 1990, A critique of pure need: An analysis of Norman Daniels's concept of health care need. International Journal of Social Economics 17: 36-45

Garber AM and CE Phelps, 1997, Economic foundations of cost-effectiveness analysis, Journal of Health Economics 16, 1-32

Gerard K and G Mooney, 1993, QALY league tables: handle with care, Health Economics 2, 59-64

Glendinning C, 1992, The Costs of Informal Care (HMSO Books, London)

Gold MR, JE Siegel, LB Russell and MC Weinstein (eds), 1996, Cost-effectiveness in health and medicine (Oxford University Press, Oxford)

Gyrd-Hansen D and J Søgaard, 1998, Discounting life-years: whither time preference?, Health Economics 7, 121-127

Hanning M, 1996, Maximum waiting-time guarantee: an attempt to reduce waiting lists in Sweden, Health Policy 36, 17-35

Harris J, 1987, QALYfying the value of life, Journal of Medical Ethics 13, 117-123

Hartunian N, N Smart and M Thompson, 1981, The incidence and economic costs of major health impairments (Lexington Books, Lexington)

Harvey CM, 1994, The reasonableness of non-constant discounting, Journal of Public Economics 53, 31-51

Hermans HEGM and WBF Brouwer, 1998a, Employeeclinics lead to discrimination in health care: myth or reality? (in Dutch), Nederlands Juristenblad 73, 1095-1101

Hermans HEGM and WBF Brouwer, 1998b, Reply (in Dutch), Nederlands Juristenblad 73, 1823-1824

Hout BA van, 1998, Discounting costs and effects differently: a reconsideration, Health Economics 7, 581-594

Hurley J, 1998, Welfarism, Extra-Welfarism and Evaluative Economic Analysis in the Health Care Sector, in: ML Barer, TE Getzen and GL Stoddard (eds), Health, Health Care and Health Economics: Perspectives on Distribution (John Wiley & Sons Ltd, Chichester)

Johannesson M, 1997, Avoiding double-counting in pharmacoeconomic studies. Pharmacoeconomics 11, 385-388.

Johannesson M and G Karlsson, 1997, The friction cost method: a comment, Journal of Health Economics 16, 249-256

Johannesson M and PO Johansson, 1996, The discounting of life-years saved in future generations — some empirical results, Health Economics 5, 329-332

Johannesson M, B Jönssen and G Karlsson, 1996, Outcome measurement in economic evaluation, Health Economics 5, 279-296

Johannesson M and U-G Gerdtham, 1996, A note on the estimation of the equity-efficiency trade-off, Journal of Health Economics 15, 359-368

Johannesson M, 1999, On aggregating QALYs: a comment on Dolan, Journal of Health Economics 18: 381-386

Jones-Lee MW, 1976, The value of life: an economic analysis (London: Martin Robertson)

Keeler EB and S Cretin, 1983, Discounting of life-saving and other nonmonetary effects, Management Science 29, 300-306

Kinney J and MAP Stephens, 1989, Caregiver Hassles Scale: Assessing the daily hassles of caring for a family member with dementia, The Gerontologist 29, 328-332

Koning J de and FAWM Tuyl. The relation between labor time, production and employment (in Dutch). Netherlands Economic Institute. Rotterdam, 1984

Koopmanschap MA, 1994, Complementary analyses in economic evaluation of health care. (Thesis, Rotterdam)

Koopmanschap MA, GJ van Oortmarssen, HME van Agt, M van Ballegooijen, JDF Habbema and JThN Lubbe, 1990, Cervical cancer screening: attendance and cost-effectiveness, International Journal of Cancer 45, 410-415

Koopmanschap MA and BM van Ineveld, 1992, Towards a new approach for estimating indirect costs of disease, Social Science and Medicine 34, 1005-1010

Koopmanschap MA and FFH Rutten, 1993, Indirect costs in economic studies: confronting the confusion, PharmacoEconomics 4, 446-454

Koopmanschap MA, FFH Rutten, BM van Ineveld and L van Roijen, 1995, The friction cost method for measuring indirect costs of disease, Journal of Health Economics 14, 171-189

Koopmanschap MA and FFH Rutten, 1996a, A practical guide for calculating indirect costs of disease, PharmacoEconomics, 10, 460-466

Koopmanschap MA and FFH Rutten, 1996b, Indirect costs: the consequences of production loss or increased costs of production. Medical Care 34, DS59-DS68

Koopmanschap MA, FFH Rutten, BM van Ineveld and L van Roijen, 1997, Reply to Johannesson's and Karlsson's comment, Journal of Health Economics 16, 257-259

Krabbe PFM, ML Essink-Bot and GJ Bonsel, 1997, The comparability and reliability of five health state valuation methods, Social Science and Medicine 45, 1641-1652

Krahn M and A Gafni, 1993, Discounting in the economic evaluation of health care interventions, Medical Care 31, 403-418

Layard R. and A Walters, 1978, Microeconomic Theory (McGraw Hill, New York)

Levin LA and B Jonsson, 1992, Cost-effectiveness of thrombolysis -- a randomized study of rt-PA in suspected myocardial infarction, European Heart Journal 13, 2-8

Lewis PA and M Charney, 1989, Which of two individuals do you treat when only their ages are different and you can't treat both? Journal of Medical Ethics 15: 28-32

Libscomb J, 1989, Time preference for health in cost-effectiveness analysis, Medical care 27, S233-S253

Liljas B, 1998, How to calculate indirect costs in economic evaluations, PharmacoEconomics 13, 1-8

Lindgren B, 1981, Costs of illness in Sweden 1964-1975 (Liber, Lund)

Loomes G and L McKenzie, 1989, The scope and limitations of QALY-measures, in: S Baldwin, C Godfrey and C Propper (eds), The quality of life: perspectives and policies. (London, Routledge)

McGuire A, J Henderson and G Mooney, 1988, The economics of health care, an introductory text (Routledge & Kegan Paul, London)

Meltzer D, 1997, Accounting for future costs in medical cost-effectiveness analysis, Journal of Health Economics 16, 33-64

Mehrez A and A Gafni, 1989, Quality-adjusted life-years, utility theory and healthy-years equivalents, Medical Decision Making 9: 142-149

Mohide EA, GW Torrance, DL Streiner, DM Pringle and R Gilbert, 1988, Measuring the wellbeing of family caregivers using the time trade-off technique, Journal of clinical epidemiology 41, 475-482

Murray CJL, 1994, Quantifying the burden of disease: the technical basis for disability adjusted life-years, Bulletin of the World Health Organization 72, 429-445

Nederlands Ziekenhuis instituut, 1997, Health care in count 5 (in Dutch), (Utrecht, Nederlandse Zorgfederatie)

Neumann J von and O Morgenstern, 1944, Theory of games and economic behaviour (Princeton University Press, Princeton NJ)

Newton JN, J Henderson and MJ Goldacre, 1995, Waiting list dynamics and the impact of earmarked funding, British Medical Journal 311, 783-784

Nietzsche F, 1885, Jenseits von Gut und Böse, Vorspiel einer Philosophie der Zukunft.

Nord E, 1990, Reducing sick leave costs by shortening waiting periods for elective surgery. Medical Decision Making 10, 95-101

Olsen JA, 1993, On what basis should be discounted? Journal of Health Economics 12, 39-54

Olsson G, LA Levin and N Rehnqvist, 1987, Economic consequences of postinfarction prophylaxis with 6 blockers: cost effectiveness of metoprolol, British Medical Journal 294, 339-342

Oosterling-Schiereck HE and R Dillmann, 1997, Nobody improves by introducing employeeclinics (in Dutch), Economisch Statistische Berichten 82: 152-153

Osterhaus JT, DL Gutterman and JR Plachetka, 1992, Health care resource and lost labor costs of migraine headache in the US, PharmacoEconomics 2, 67-76

Parker G, 1990, Spouse Carers. Whose quality of life?, in: S Baldwin, C Godfrey and C Propper (eds), The quality of life: perspectives and policies. (London, Routledge)

Parsonage M and H Neuburger, 1992, Discounting and health benefits, Health Economics 1, 71-76

Pauly MV, 1995, Valuing health care benefits in money terms, in: F. Sloan (ed), Valuing health care (Cambridge University Press, Cambridge)

Pauly MV, 1997, Who was that straw man anyway? A comment on Evans and Rice, Journal of Health Politics, Policy and Law 22, 467-473

Pigou AC, 1932, The economics of welfare (4th edition, London, Macmillian)

Pliskin JS, DS Shepard and MC Weinstein, 1980, Utility functions for life years and health status, Operations Research 28, 206-224

Posnett J and Jan S, 1996, Indirect cost in economic evaluation: the opportunity cost of unpaid input, Health Economics 5, 13-23

Pratt JW and RJ Zeckhauser, 1996, Willingness to pay and the distribution of risk and wealth, Journal of political economy 104, 747-763

Rawls J, 1971, A theory of justice (Harvard University Press, Cambridge, MA)

Reid MG, 1934, Economics of household production (Wiley, New York)

Rice T, 1997, Can markets give us the health system we want?, Journal of Health Politics, Policy and Law 22, 383-426

Roijen L van, L Essink-Bot, MA Koopmanschap, BC Michel and FFH Rutten, 1995a, Societal Perspective on the Burden of Migraine in the Netherlands, PharmacoEconomics 7, 170-179

Roijen L van, MA Koopmanschap, FFH Rutten and PJ van der Maas, 1995b, Indirect costs of disease; an international comparison, Health policy 33, 15-29

Roijen L van, ML Essink-Bot, MA Koopmanschap, Michel BC, Rutten FFH, 1996, Labor and health status in economic evaluation of health care, International Journal of Technology Assessment in Health Care 12, 405-415

Russell LB, 1986, Is prevention better than cure? (The Brookings Institution, Washington DC)

Russell LB, MR Gold, JE Siegel, N Daniels and MC Weinstein, 1996, The Role of Cost-effectiveness Analysis in Health and Medicine, JAMA 276, 1172 - 1177

Rutten FFH, BM van Ineveld, R van Ommen, BA van Hout and R Huijsman, 1993, Costcalculation in health care: practical guiselines (in Dutch), (Van Arkel, Utrecht)

Seeverens HJJ, 1997, Employeeclinics and justice (in Dutch), Economisch Statistische Berichten 82, 151-152

Sen A, 1970, Collective choice and Social Welfare (San Fransisco, Holden-Day)

Sen, A, 1977a, Social choice theory: a re-examination, Econometrica 45, 53-90

Sen, A., 1977b, On weights and measures, Econometrica 44, 1539-1572

Sen, A., 1980, Equality of what?, in: The Tanner Lectures on Human values (Cambridge University Press, Cambridge)

Sen, A., 1995, Rationality and Social Choice, The American Economic Review 85, 1-24

Severens JL, RJF Laheij, JBM Jansen, EH van der Lisdonk and ALM Verbeek, 1998, Estimating the cost of lost productivity in dyspepsia, Alimentary pharmacology and therapeutics 12, 919-923

Smith K and K Wright, 1994, Informal care and economic appraisal: A discussion of possible methodological approaches, Health Economics 3: 137-148

Swinkels H, 1996, Waiting for medical treatment: waitinglists (in Dutch), Maandbericht gezondheidstatistiek, CBS, Oktober

Torrance GW, 1986, Measurement of health state utilities for economic appraisal: a review, Journal of Health Economics 5, 1-30

Viscusi WK, 1979, Employment Hazards: an investigation of market performance (Cambridge: Harvard University Press)

Wagstaff A, 1991, QALYs and the equity-efficiency trade-off, Journal of Health Economics 10: 21-41

Waugh N and D Scott, 1998, Letter to the Editor: How should different life-expectancies be valued?, British Medical Journal 316, 1316

Weinstein MC and WB Stason, 1976, Hypertension: A policy perspective (Cambridge, Harvard University Press)

Weinstein MC and WB Stason, 1977, Foundations of cost-effectiveness analysis for health and medical practices, New England Journal of Medicine 296, 716-721

Weinstein MC, 1986, Risky choices in medical decision: a survey. Geneva Papers on Risk and Insurance 11, 197-216

Weinstein MC, JE Siegel, MR Gold, MS Kamlet and LB Russell, 1996, Recommendations of the panel on cost-effectiveness in health and medicine, Journal of the American Medical Association 276, 1253-1258

Weinstein MC and WG Manning Jr., 1997, Theoretical issues in cost-effectiveness analysis, Journal of Health Economics 16, 121-128

Weinstein MC, JE Siegel, AM Garber, J Libscomb, BR Luce, WG Manning Jr. and GW Torrance, 1997, Productivity costs, time costs and health related quality of life: a response to the Erasmus Group, Health Economics 6, 505-510

Willems JHBM, 1998, Privatisation in social security (in Dutch), Medisch Contact 53, 423-426

Williams A, 1985, Economics of coronary artery bypass grafting, British Medical Journal 291, 326-329

Williams A, 1988, Economics and the rational use of medical technology, in: FFH Rutten and SJ Reiser (eds), The economics of medical technology (Berlin, Springer Verlag)

Williams A, 1993, Cost-Benefit analysis: applied welfare economics or general decision aid, in: A Williams and E Giardina (eds), Efficiency in the public sector (Edward Elgar, London)

Williams A, 1997, Intergenerational equity: an exploration of the 'fair innings' argument, Health Economics 6, 117-132

Wright K, 1987, The economics of informal care. Centre for Health Economics, Discussion Paper 23 (University of York, York)

Zarit SH, KE Reever and J Bach-Peterson, 1980, Relatives of the impaired elderly: Correlates of feelings of burden, The Gerontologist 20, 649-655.

Samenvatting

Tijd en tijdskosten in economische evaluaties vanuit een maatschappelijk perspectief

Inleiding

Economische evaluaties worden in toenemende mate gebruikt om beleidsmakers te ondersteunen bij de besluitvorming met betrekking tot de implementatie en vergoeding van gezondheidszorgprogramma's. Verschillende evaluatietechnieken kunnen hierbij worden gebruikt, zoals kostenminimalisatie, kosten-consequentie analyse en kosten-baten analyse. Een andere, veelvuldig gebruikte evaluatie vorm is de kosten-effectiviteitsanalyse (of kosten-utiliteitsanalyse), waarbij de gezondheidseffecten van een interventie worden afgewogen tegen de kosten ervan. Vaak wordt als uitkomstmaat zogenaamde QALYs (*quality adjusted life years*) gebruikt, voor kwaliteit gecorrigeerde levensjaren. Deze uitkomstmaat maakt het mogelijk om de effecten van verschillende interventies, eventueel gericht op verschillende gezondheidsproblemen, met elkaar te vergelijken. Met behulp van kosten-effectiviteitsanalyses kunnen idealiter uiteindelijk met het gezondheidszorgbudget die interventies worden uitgevoerd die de meeste gezondheidswinst opleveren per kosten-eenheid. Het globale doel van economische evaluaties wordt daarom veelal gezien als het maximaliseren van gezondheidseffecten met een gegeven budget.

Over het algemeen wordt als meest geschikte perspectief voor het uitvoeren van een kosteneffectiviteitsanalyse het maatschappelijke perspectief aangemerkt. Dit houdt in, dat alle kosten en effecten in principe moeten worden meegenomen in een analyse, ongeacht wie de kosten draagt of wie de effecten ondervindt. Deze brede maatschappelijke

visie op kosten-effectiviteitsanalyse heeft gevolgen, zowel voor de bepaling van kosten als voor de bepaling van effecten. Zo kan men zich niet enkel beperken tot de direct medische kosten, maar dienen tevens andere maatschappelijke kosten te worden meegenomen. Het consistent doorvoeren van het maatschappelijke perspectief in alle onderdelen van de economische evaluatie blijkt niet eenvoudig. Bijvoorbeeld bij het bepalen van productiviteitskosten is het gemakkelijk om 'terug te vallen' op een individualistisch perspectief, zoals in de veelgebruikte human capital methode. Het volgen van een maatschappelijk perspectief impliceert volgens sommigen ook de noodzaak om kosten en effecten op een eerlijke wijze weer te geven, ongeacht wie ze ervaart (Gold et al., 1996). Hierbij kan bijvoorbeeld gebruik worden gemaakt van rechtvaardigheidstheorieën zoals die ontwikkeld door Rawls (1971).

Uiteraard is voor de vergelijkbaarheid van resultaten van verschillende analyses een uniforme methodologie voor kosten-effectiviteitsanalyses van belang. Echter, op verschillende onderdelen bij de bepaling van kosten en effecten ontbreekt een duidelijke consensus. Deels kan het ontbreken van consensus over belangrijke methodologische vraagstukken worden teruggevoerd op een verschil in generale opvatting over kosten-effectiviteitsanalyse. Twee brede stromen kunnen worden onderscheiden: de welvaartseconomische benadering (*welfarism*) en de beleidsmaker's benadering (*decision-maker's approach*). In de eerste visie ziet men kosten-effectiviteitsanalyse als een toepassing van welvaartseconomische theorie. De tweede visie is breder en pragmatischer. Hierin wordt economische evaluatie gezien als een hulpmiddel voor beleidsmakers, zonder binnen de strikte welvaartseconomische theorie te blijven. De keuze voor een van beide stromen heeft ook consequenties voor de methodologische keuzen in een economische evaluatie. In dit proefschrift wordt uitgegaan van de tweede visie op kosten-effectiviteitsanalyse.

Onderzoeksvragen

Naast een uiteenzetting van welvaartseconomische benadering en de beleidsmaker's benadering en de methodologische consequenties van het kiezen van een van beide visies worden in dit proefschrift enkele controversiële methodologische keuzen binnen kosten-effectiviteitsanalyses besproken en worden aanbevelingen gedaan over de te gebruiken methoden. De nadruk ligt daarbij op de waardering van tijd

van patiënten wanneer zij als gevolg van ziekte en/of behandeling hun normale tijdsbesteding, betaald en onbetaald werk en vrije tijd, niet kunnen continueren. De waardering van tijd van informele zorgverleners wordt tevens behandeld, aangezien zij normale tijdsbestedingen opgeven om informele zorg te kunnen verlenen. Daarnaast wordt het disconteren van gezondheidseffecten die later in de tijd optreden in economische evaluaties besproken. In het kader van waardering van patiëntentijd en verschillen in de waardering van deze tijd over verschillende groepen in de maatschappij wordt ook bezien in hoeverre versnelde hulp aan werknemers, om productieverliezen te vermijden, op een efficiënte en rechtvaardige wijze kan worden verleend.

Belangrijke vragen die in dit proefschrift aan de orde komen zijn:

1. Hoe kan tijd van patiënten worden gewaardeerd in een economische evaluatie vanuit een maatschappelijk perspectief en welke onderdelen dienen daarbij in termen van kosten en welke in termen van effecten te worden meegenomen?
2. Kan, vanuit een maatschappelijk perspectief, versnelde hulp aan werkenden op een efficiënt en rechtvaardige wijze worden georganiseerd?
3. Hoe kan de tijd van informele zorgverleners worden gewaardeerd in economische evaluaties vanuit een maatschappelijk perspectief?
4. Moeten toekomstige gezondheidseffecten in een economische evaluatie worden gedisconteerd en zo ja, hoe?

Deze vragen zijn alle gerelateerd aan het centrale thema van dit proefschrift:

Hoe verandert het hanteren van een maatschappelijk perspectief, vanuit een beleidsmaker's benadering, de methodologie van economische evaluaties, met name die methoden met betrekking tot tijd en tijdskosten.

Twee visies

Alhoewel kosten-effectiviteitsanalyse zowel in de welvaartseconomische benadering als in de beleidsmaker's benadering er op is gericht beleidsmakers te ondersteunen bij besluitvorming, leiden de verschillen

tussen beide visies tot verschillen in aanbevolen methoden op verschillende gebieden zoals het meenemen van rechtvaardigheidsaspecten, de waardering van productieverliezen, de interpretatie van QALYs, enzovoort (hoofdstuk 2). Het gebruik van vaak individualistische, welvaartstheoretische nutsmodellen als basis voor methodologische aanbevelingen lijkt niet altijd te verenigen met de keuze voor een maatschappelijk perspectief. Ook kunnen in dergelijke modellen elementen ontbreken, die in werkelijke besluitvormingsprocessen van belang zijn, zoals rechtvaardigheidsaspecten.

Alhoewel binnen de welvaartseconomie nut de belangrijkste uitkomstmaat is, kan dit binnen gezondheidseconomisch onderzoek als minder geschikt worden aangemerkt. Binnen de gezondheidszorg ligt de nadruk wellicht meer op het produceren van gezondheid, het reduceren van ongezondheid en het herstellen van fysieke en mentale *mogelijkheden* die gezondheid met zich meebrengt (Culyer, 1990). In dat opzicht kan het doel van de beleidsmaker's benadering worden omschreven als het optimaliseren van (gewogen) gezondheidseffecten met een gegeven budget.

Tijd van patiënten

De waardering van veranderingen in de tijdsbesteding van patiënten is een gebied waarin het hanteren van het maatschappelijk perspectief duidelijke consequenties heeft voor de gebruikte (waarderings-)methoden in economische evaluaties. De human capital methode kan worden beschouwd als een waarderingsmethode vanuit een individualistisch perspectief. In dat opzicht is de methode minder geschikt om te gebruiken in een analyse vanuit een maatschappelijk perspectief, waarbij normale vervangingsmechanismen in de analyse en de waardering dienen te worden betrokken. De frictiekosten methoden is momenteel de enige praktische methode die hier rekening mee houdt en in dat opzicht start vanuit een maatschappelijk perspectief.

Een belangrijk aspect bij de waardering van patiëntentijd is de complementariteit van kwaliteit van leven meting en kostenmeting. Niet alleen moet men keuzes maken met betrekking tot welke tijdselementen in termen van kosten en welke in termen van kwaliteit van leven veranderingen moeten worden aangemerkt, maar ook moet worden

bezien in hoeverre bestaande kwaliteit van leven instrumenten reeds bepaalde elementen van patiënten tijd beslaan. Dit voorkomt het buiten beschouwing laten of juist dubbeltellen van tijdskosten. Verminderd rolfunktioneren en veranderingen in de mogelijkheden om normale vrijetijdsbestedingen uit te voeren kunnen bij voorkeur in termen van kwaliteit van leven worden gewaardeerd. Verloren productie bij betaald en onbetaald werk kan bij voorkeur worden gewaardeerd in termen van kosten (hoofdstuk 3). De precieze complementariteit tussen kosten en kwaliteit van levenwaardering verdient nadere bestudering, met name de mogelijke spanning tussen kwaliteit van leven waardering door het generale publiek en meer individualistische kostengegevens.

Alhoewel er een relatie bestaat tussen kwaliteit van leven en tijdskosten is het niet aan te bevelen om productiviteitskosten, i.c. de kosten van verloren productie, in termen van kwaliteit van leven te waarderen, zoals recent voorgesteld door een panel van vooraanstaande Amerikaanse gezondheidseconomen (Gold et al., 1996). Er bestaan belangrijke bezwaren tegen een dergelijke waarderingswijze, zoals uiteengezet in hoofdstukken 4 en 5. Een belangrijk bezwaar is dat er enkel een zwakke relatie bestaat tussen inkomen en productiviteit in veel ontwikkelde landen. Sociale voorzieningen zullen het inkomen van personen die wegens ziekte niet langer (betaald) productief kunnen zijn op een bepaald niveau handhaven. Ook lijkt het op deze wijze waarderen van productiviteitskosten een meer individueel perspectief te hanteren en geen maatschappelijk perspectief.

Het meten en waarderen van verminderde productiviteit bij betaald werk zonder verzuim behoeft meetinstrumenten die op een accurate wijze het prestatieniveau van personen kan meten op dagen dat zij bij het werk gehinderd worden door gezondheidsproblemen. Zoals uiteengezet in hoofdstuk 6, lopen de schattingen van de kosten van dit type productiviteitskosten uiteen wanneer men verschillende meetinstrumenten gebruikt. Ziekte op het werk blijkt in de in hoofdstuk 6 beschreven populatie vrij frequent voor te komen. Gemiddeld gaf 7% van de respondenten op een willekeurige werkdag aan dat zij gezondheidsproblemen ondervonden terwijl zij werkten. Wel bleek dat het prestatieniveau van deze personen in vergelijking tot normaal niet erg daalde, waardoor de gerelateerde productiviteitskosten op bedrijfsniveau

relatief laag zijn. Echter, voor specifieke patiëntengroepen kunnen deze kosten aanzienlijk zijn. Het inschatten van de kwaliteit van het geleverde werk in vergelijking tot normaal wordt veelal buiten beschouwing gelaten, maar kan een interessant aspect van geleverde arbeid toevoegen. De hiertoe ontwikkelde QQ-methode zou in de toekomst verder kunnen worden toegepast en gevalideerd, samen met de andere gangbare meetmethoden. Ook specifieke vervangingsmechanismen door collega's en de invloed van de organisatie van arbeid in de onderzoeksomgeving zouden profiteren van verder onderzoek.

Informele zorg

Informele zorgverleners geven hun normale tijdsbesteding (deels) op om zorg te kunnen verlenen wanneer in hun directe (sociale) omgeving ziekte optreedt. De waardering van informele zorg kan van invloed zijn op de resultaten bij bepaalde gezondheidszorginterventies, bijvoorbeeld bij zorg voor Alzheimerpatiënten (Busschbach et al., 1998). Methoden die trachten de volledige maatschappelijke impact van informele zorg te waarderen zijn daarom belangrijk. In hoofdstuk 7 is een alternatieve tijdwaarderingsmethode besproken, waarbij naast kosten ook de kwaliteit van leven verandering als gevolg van het geven van informele zorg wordt meegenomen. Indien vanuit een maatschappelijk perspectief alle kosten en effecten dienen te worden meegenomen, *ongeacht wie ze ondervindt,* dan is ook een kwaliteit van leven verandering bij informele zorgverleners in principe relevant. Met een combinatie van kwaliteit van leven en kosten kan op deze manier de volledige veranderde tijdsbesteding van informele zorgverleners worden gewaardeerd. De resultaten van kwaliteit van levenmeting bij informele zorgverleners dienen afzonderlijk van die bij patiënten te worden gerapporteerd.

Versnelde hulp voor werkenden

Het hanteren van een maatschappelijk perspectief kan tevens misallocaties van middelen aantonen tussen gezondheidszorg en andere maatschappelijke sectoren. Het rantsoeneren van gezondheidszorg, leidend tot lange wachtlijsten in de zorg, draagt weliswaar bij tot het beperken van uitgaven in de gezondheidszorg, maar ook tot hoge kosten in termen van productieverliezen of hogere productiekosten. Indien een dergelijke relatie is ontdekt kan worden bezien of een betere allocatie van

middelen mogelijk is, waar alle partijen van kunnen profiteren. Wanneer werknemers als gevolg van ziekte niet langer kunnen functioneren in hun werk, kan dit aanzienlijke maatschappelijke kosten teweegbrengen. Werkgevers zijn in Nederland sinds 1994 gedeeltelijk en sinds 1996 geheel verantwoordelijk voor de loondoorbetaling bij ziekte gedurende het eerste jaar. Daardoor worden zij sinds die tijd rechtstreeks geconfronteerd met de kosten die absentie teweeg kan brengen en zijn zij zich bewuster geworden van het bestaan van deze kosten. Een van de consequenties van de verschuiving van deze kosten richting de individuele werkgevers, is de discussie over versnelde hulp aan werknemers in zogenaamde bedrijvenpoli's. Zoals besproken in hoofdstuk 8, kunnen deze poli's in principe zo worden georganiseerd, dat versnelde hulp voor werknemers zowel efficiënt als rechtvaardig kan worden genoemd. Hiervoor is het met name noodzakelijk dat voorheen onbenutte capaciteit (bijvoorbeeld weekeinden en avonduren) voor deze versnelde hulp wordt benut. Daardoor kan de druk binnen de reguliere openingsuren verminderen, waardoor ook niet-werknemers sneller dan voorheen geholpen kunnen worden. Belangrijk is het ook te beseffen dat versnelde hulp kopen in andere Europese landen wellicht niet valt tegen te gaan in het licht van recente Europees rechterlijke uitspraken (Hermans en Brouwer; 1998a; Hermans en Brouwer, 1998b; Brouwer en Hermans, 1999).

Het is overigens interessant te zien, dat een expliciete beschouwing van productiviteitskosten in het kader van bedrijvenpoli's en het verbinden van consequenties daaraan met betrekking tot prioritering in de zorg, heeft geleid tot een grote mate van maatschappelijk verzet tegen voorrang of tweedeling in de zorg. Immers, bij de beoordeling van verschillende gezondheidszorginterventies met behulp van economische evaluaties worden deze kosten ook meegenomen en kunnen deze kosten soms van doorslaggevend belang zijn (Koopmanschap, 1994). Zo worden op het niveau van een enkele interventie of evaluatie deze maatschappelijke kosten meegerekend, hetgeen kan leiden tot een bevoordeling van betaald werkenden op interventieniveau, terwijl een explicietere weging van deze kosten, zoals in het geval bij versnelde hulparrangementen voor werknemers, niet geoorloofd wordt geacht. Het verdient aanbeveling instrumenten te ontwikkelen, die inzicht kunnen verschaffen in de verdelende effecten van het meenemen van productiviteitskosten in economische evaluaties.

Disconteren van effecten

Disconteren van gezondheidseffecten omdat zij later in de tijd optreden is een veelbesproken onderwerp binnen de gezondheidseconomische literatuur. Veelal wordt nu gekozen voor een disconteringspercentage, gelijk aan die van kosten (vaak tussen de drie en de vijf procent). De keuze voor eenzelfde disconteringspercentage voor kosten en effecten wordt veelal gebaseerd op het consistentieprincipe (Weinstein en Stason, 1977) en de uitstelparadox (Keeler en Cretin, 1983). Echter, zoals uiteengezet in hoofdstuk 9, kunnen er vraagtekens worden gezet bij het gebruik van eenzelfde disconteringspercentage voor kosten en effecten en de vaak geclaimde noodzaak daarvan. Ook voor het disconteren van effecten kan het innemen van een maatschappelijk perspectief consequenties hebben. Met name de eerlijke representatie van toekomstige effecten en effecten in toekomstige generaties speelt hierbij een rol. Alhoewel het disconteren van effecten vanuit een individueel perspectief wellicht begrijpelijk is, lijkt dit minder het geval vanuit het maatschappelijke perspectief. Gebruik makend van de rechtvaardigheidstheorie van Rawls (1971), kan men beargumenteren dat een meer 'tijdloos' of intergenerationeel maatschappelijk perspectief ingenomen moet worden, waarin de basis voor het disconteren van effecten daadwerkelijke verschillen tussen personen over tijd is. Veranderende (voor kwaliteit gecorrigeerde) levensverwachting kan een dergelijk objectief en relevant verschil tussen personen over de tijd zijn. Samen met afnemend marginaal grensnut van additionele QALYs kan dit een objectievere basis vormen voor het disconteren van effecten dan de huidige wijze van disconteren. Het blijft echter op dit moment moeilijk een specifiek disconteringspercentage voor effecten aan te raden, omdat het (nog) onduidelijk is hoe de waardering van additionele effecten over de tijd verandert wanneer de levensverwachting verandert. Derhalve lijkt het raadzaam het disconteren van effecten achterwege te laten en daarbij te vermelden wanneer de effecten optreden. De keuze voor een eventuele weging van toekomstige effecten berust dan bij de beleidsmaker, die daarbij ook overwegingen van rechtvaardigheid en de belangen van huidige en toekomstige generaties in ogenschouw kan nemen.

Epiloog

Het hanteren van een maatschappelijk perspectief in economische evaluaties van gezondheidszorgprogramma's heeft consequenties voor de methodologie in deze evaluaties. In dit proefschrift zijn enkele van deze consequenties, voor de waardering van tijd van patiënten en informele zorgverleners en het disconteren van gezondheidseffecten besproken vanuit een beleidsmaker's benadering van economische evaluatie. Verschillende methoden zijn besproken en onderzocht op hun geschiktheid te worden gebruikt in evaluaties vanuit het maatschappelijk perspectief. Ook zijn alternatieve waarderingsmethoden aangedragen.

Verdere harmonisatie van gebruikte methoden in economische evaluaties en de verdere ontwikkeling van een uniforme presentatiewijze van de resultaten zijn twee grote uitdagingen voor de toekomst. Het eerste kan leiden tot een betere vergelijkbaarheid tussen verschillende studies en meer transparantie in hoe de resultaten tot stand zijn gekomen. Het tweede kan leiden tot een betere informatieoverdracht naar de beleidsmakers. Het ontwikkelen van een volledige 'gereedschapskist' voor het uitvoeren van economische evaluaties vanuit een maatschappelijk perspectief zal veel inspanning vergen, maar is van groot belang. Dit proefschrift hoopt iets aan deze ontwikkeling bij te dragen.

Curriculum vitae

Werner Brouwer was born in Stadskanaal in 1972. After graduating from the Ichthus College in Veenendaal in 1990, he studied Economics at the Erasmus University in Rotterdam. During his study he assisted in research projects of the *Stichting Economisch Onderzoek Rotterdam* (SEOR) and also in teaching micro-economics.

After graduating in 1996, he started working at the Department of Health Policy and Management and the institute for Medical Technology Assessment (iMTA) of the Erasmus University Rotterdam. Besides his PhD-project he has been involved in several iMTA-projects and teaching at the Department of Health Policy and Management. Since 1998, he is editor of the VGE-Infobulletin, the quarterly magazine of the Dutch/Flemish Association for Health Economics.

Acknowledgements

During the time I studied economics I never imagined that I would prolong my stay at the Erasmus University in Rotterdam. The fact that I did and that I have not regretted it since, is largely due to the stimulation of several persons whom I would like to thank here.

I wish to thank Marc Koopmanschap and Frans Rutten. They not only took the chance of hiring me (an unsolved mystery), but also provided me with the freedom of pursuing my interests and ideas. Thanks for your enthusiastic stimulation over the years. Marc, as my co-promotor, was always in for new brainstorm-sessions, although the always mentioned whiskey and cigars never became reality!

Special thanks to Jaap van Dijk. During my study of economics, Jaap was the one person who stimulated me to continue and aroused my curiosity for micro-economics. Although he was my teacher at the start, we became friends during my final year of economics and remained friends afterwards. He stimulated my choice for taking the position at the Department of Health Policy and Management and was always around when I had a question or idea. My dear friend Jaap died 31 August 1999. Rotterdam isn't the same without you.

Furthermore, all the colleagues and friends at the Department of Health Policy and Management and the institute for Medical Technology Assessment with whom I had the pleasure of working with. In the context of this thesis I especially need to thank co-authors Erik Schut and Ben van Hout. I am grateful to Fern Terris-Prestholt for correcting my English in various chapters. Reinhilde Klostermann and Ad Vermaas shared the pain of sharing a room with me. Thanks for putting up with me!

A word of thanks to my family and friends, for their support and enthusiasm. Job, I hope we'll have more time in the future to 'lunz'

together. You have been a great friend through the years. My parents I wish to thank for their continuous support and interest. I am very lucky to have you both!

Finally, I want to thank Janine. You have accomplished something incomparably more important this year than I did by finishing this thesis: giving birth to our beautiful son Niels. My love for you both is timeless.