Cost-utility Analysis

JND de Neeling

Cost-utility Analysis

JND de Neeling

No. A03/01E, The Hague, February 3, 2003

The Health Council of the Netherlands, established in 1902, is an independent scientific advisory body. Its remit is "to advise the government and Parliament on the current level of knowledge with respect to public health issues..." (Section 21, Health Act).

The Health Council receives most requests for advice from the Ministers of Health, Welfare & Sport, Housing, Spatial Planning & the Environment, Social Affairs & Employment, and Agriculture, Nature & Food Quality. The Council can publish advisory reports on its own initiative. It usually does this in order to ask attention for developments or trends that are thought to be relevant to government policy.

Most Health Council reports are prepared by multidisciplinary committees of Dutch or, sometimes, foreign experts, appointed in a personal capacity. The reports are available to the public.

The Health Council of the Netherlands is a member of INAHTA, the international network of health technology assessment (HTA) agencies that promotes and facilitates information exchange and collaboration among HTA agencies.

This report can be downloaded from www.healthcouncil.nl.

Preferred citation:
JND de Neeling. Cost-utility Analysis. The Hague: Health Council of the Netherlands, 2003; publication no. A03/01E.

ISBN: 90-5549-556-5

Contents

Foreword

This report has been written by JND de Neeling, PhD, secretary to the Health Council, in a personal capacity, as a background paper for the Committee on Contours of the Basic Health Care Benefit Package. The conclusions that the committee has drawn are incorporated into the advisory report itself.

It has become a well thought-out and absorbing report. It sketches the possibilities and limitations of cost-utility analysis when mapping the efficiency of health care. The report makes it clear that the cost-utility analysis methodology is fraught with a number of problems, so that it is only modestly useful in comparisons of the cost-utility of different facilities across the entire breadth of health care. On the one hand, the analysis presented thus embodies challenges for the scientific community in the field of health care efficiency analysis. On the other hand, it offers a realistic view on the potential role of cost-utility analysis within a policy framework.

(signed),
Professor JA Knottnerus
President of the Health Council of the Netherlands

Abstract

In defining a collectively financed basic health care benefit package, can cost-utility analysis (CUA) serve as an instrument to measure the efficiency of all preventive and curative health care services in a uniform way? This question, answered affirmatively by the Scientific Council for Government Policy in its 1997 report 'Public Health Care', lies at the heart of this background study, which is written in support of the discussion in the Health Council Committee on Contours of the Basic Health Care Benefit Package. The study presents an inventory of the assumptions and choices in CUA methodology, and provides an overview of what has been written about them in the health economic, medical and medical ethics literature.

First there is a consideration of how the general guiding notion of efficiency is transformed in CUA into a number, applicable to the entire field of health care. This so-called cost-utility ratio represents the input ('costs') divided by the output ('health gain') of health care services.

The subsequent and most comprehensive chapter is devoted to the generally applicable measure for health gain that distinguishes CUA from other forms of efficiency analysis: the Quality Adjusted Life Year (QALY). The road to the QALY appears to be paved with a series of pitfalls and traps that are still the topic of much discussion. These include problems in integrating life's 'quality' and 'quantity', in tallying up the QALYs of different people and in describing health states in a standardized manner. There are also problems in formulating a theory about health state preferences, in developing procedures to measure those preferences and in determining whose preferences will be used.

The following two chapters consider aspects of CUA that are inherent to every form of quantitative efficiency analysis: quantifying the costs and converting future health gain to its present value ('discounting'). The scientific literature shows that there appear to be differing approaches to different types of costs (for example, of the direct health care costs, the productivity costs and the costs in added years of life) that may lead to important differences in cost estimates. There is an equal lack of consensus on the question as to whether, and if so how, future health gains should be discounted.

Subsequently, it is observed that CUA, applied as an aid to achieving a maximum number of QALYs gained with a given health care budget, has a host of implications for the distribution of the health gain attained.

The final chapter concludes that the QALY and CUA are, more than 25 years after their introduction, still open to discussion in several respects. In assessing the efficiency of health services, CUA offers only limited footing. Alternatives to the QALY, such as the saved young life equivalent and the willingness to pay approach, deserve further investigation. But to define a collectively financed basic package, it will first of all be necessary to develop a transparent decision making procedure. This procedure must ensure that there is active participation within sectors of health care of the parties familiar with those sectors from the inside, primarily care providers and patients. CUA can be an aid towards reaching an overview and agreement in this decision making process.

Chapter 1

Introduction

1.1 Background

Governments throughout the entire western world are grappling with the challenge of how best to meet increasing demands on health care systems while also trying to limit the demands on their budgets (Kle93). This is equally the case in the Netherlands.

In February 2001, the Dutch Minister of Health, Welfare & Sport, preparing for the political discussion on reforming the health care insurance system, asked the Health Council of the Netherlands for its view on "scientifically based, practically useful criteria to define a basic package of health care services". The Minister drew particular attention to the criterion of efficiency, and indicated that she would like to be informed of the Council's advice on "the possibility of using the efficiency criterion in determining the composition of a basic package".

The Chairman of the Health Council passed the fulfilment of the Minister's advisory request to the Committee on Contours of the Basic Health Care Benefit Package. This report is a background study in support of the discussion within the committee on the usefulness of the efficiency criterion.

1.2 The Public Health Care report

In 1997 the Netherlands Scientific Council for Government Policy (*Wetenschappelijke Raad voor het Regeringsbeleid*, WWR) published the report 'Public Health Care'

(*Volksgezondheidszorg*) (WWR97). The Minister of Health, Welfare & Sport expressly asked the Health Council to include this report in its consideration.

The central problem defined in the WRR report is how to guarantee the general accessibility and a uniformly high quality of health care over the longer term for everyone in the Netherlands. The WRR determined that it is essential to define a collectively financed basic care package. The council acknowledged the basic role played by a specific operationalization of the efficiency criterion, the cost-utility ratio (C/U ratio),* in the decision making on defining the preventive and curative care package. The C/U ratio (or cost-utility) is defined as the *additional* costs divided by the *additional* output (measured by a standard measure of health gain applicable to the entire health care) from the health care service to be judged on its cost-utility in comparison to a relevant alternative.** The alternative may be usual care, a competing service or no service. The instrument to determine the cost-utility is the cost-utility analysis (CUA).

As far as defining the *preventive* services package is concerned, according to the WRR, the primary guideline is to maximize the health gain given a specific budget. This requires all preventive programmes to be ranked according to cost-utility. The WRR suggests using the Disability Adjusted Life Year (DALY, derived from the WHO Global Burden of Diseases Study (Mur96, Mur97)) as a unit of health gain for this. The DALY is a measure that resembles in many respects the Quality Adjusted Life Year (QALY), familiar from the medical and health economic literature. According to the WRR, the maximum health gain can be achieved by financing the various programmes in order of decreasing efficiency (i.e. increasing cost-utility ratios), until the available budget is exhausted (pp. 148-50).

In *curative* health care, according to the WRR, one should strive towards an "equalization of C/U ratios" across all categories of curative services (pp. 150-2). The idea is that the decision making on whether or not to finance collectively mainly new curative services must be governed by the principle that one identical maximum amount for realizing one DALY holds for all categories. The WRR believes that the application of the "methodology" of equalizing the cost-utility ratio may lead to an "objectifying" of the

---

* In its report the WRR used the general term cost-effectiveness ratio. In line with accepted health economic expression, in calculating the cost *effectiveness* ratio, the output of health care services can be expressed in various measures of effectiveness: increased life expectancy, complaint-free days, millimetres of high blood pressure reduction, etc. Specific requirements are imposed on the outcome measure when calculating the cost-*utility* ratio: it must be applicable to the entire health care and must represent the values that are accorded to the total of all the health effects of a service (Dru97).

** An efficiency analysis may also lead to the conclusion that the costs and/or the health gain of the studied service S are *lower* than those of the alternative A. When the costs of S are *lower* and the gain is *higher* than those of A, then it is stated in health economic that A is *dominated* by S. If it is the other way around, then S is dominated by A. In both cases calculating a cost-utility ratio is not meaningful. If both the costs and gain of S are *lower* than those of A, then there is the possibility of reversing the comparison to refer to the cost-utility ratio of A in comparison to S. In fact, in most instances it would be possible to refer to *additional* costs and an *additional* health gain of S in comparison to A.

decision making on whether or not to make curative services collectively available (p. 155). According to the Council, this does not alter the fact

"[...] that awkward choices still remain even in the suggested methodology. Thus, when using the DALY measure, a choice must be made between what will and will not be regarded as costs (indirect costs, direct costs etc.) and in which manner the benefits should be valued. These difficult points should certainly not be a reason not to use such a methodology. In some respects, the matter is one of resolving technical issues. Not only can these choices be expressly discussed, but the additional advantage remains that all services can be weighed in the same manner" (p. 155).

The WRR leaves unanswered in just what way the "awkward" issues can be considered technical. The suggestion is that these issues are not only technical in the sense that they require specific expertise in order to be fathomed. They are also technical because they can be resolved in different ways (so that the "machinery" of CUA can thus demonstrate varying views of parts) without this having any essential influence on the eventual result. After all, "the additional advantage remains that all services can be weighted in the same manner." The WRR does not argue why a specific collection of methodological choices (applied to various services with divergent health effects among diversely composed groups of patients) would lead to the same ranking by virtue of cost-utility as would another CUA methodology set up differently in terms of "technical" aspects.

1.3 The Limits to Health Care advisory

The advisory 'Limits to Health Care' (*Grenzen van de gezondheidszorg*) (GR86) appeared more than ten years before the Public Health Care report. In this advisory, the Health Council of the Netherlands accorded the CUA a much more modest role than the WRR would do later. According to the Health Council, CUA results should not be decisive in ranking services, and thus not in the decision making. They could be "no more than *helpful* in raising the quality of the decision making" (p. 89). A ranking according to cost-utility could result in arbitrariness. In fact, such a ranking would be dependent on the supply of health care services available at any given time and on the limited capacity to carry out CUAs. Furthermore, it might be overlooked that in the course of time, with increasing experience and wider application, the cost-utility of services could improve. Finally, there are other criteria alongside cost-utility that are relevant to the inclusion of services in the package. Is the service of life and death importance or otherwise essential? Is the service accessible to anyone who is entitled to it? Does the care contribute towards the desired distribution of health care within the population and among groups of patients (p. 90)? All these were issues posed by the Health Council in 1986.

1.4 This report

Fifteen years have elapsed since the appearance of the Health Council's Limits to Health Care advisory. Countless articles have been published discussing the principles and the methodology of cost-utility analysis and forwarding new proposals to solve persistent problems. The number of available CUAs has multiplied. It is now expedient to reassess the old Health Council view on the usefulness of cost-utility analysis in the light of these scientific developments. This is particularly so since the Minister's advice request expressly referred to the WRR report that is in conflict with the earlier viewpoint of the Health Council.

The aim of this report is to initiate the defining of a new Health Council viewpoint. Its primary goal is limited. It involves creating an inventory of the assumptions and methodological problems and the limitations of the applicability of CUA that arise from them. The report is thus to a large extent devoted to the "awkward choices" in the CUA methodology that the WRR has already mentioned. It provides an overview of facts, opinions and considerations with respect to these choices, that are highlighted in the health economic, medical and medical ethics literature. Following from this, the final chapter explores what lessons may be learnt from this for the broader question of whether the efficiency criterion can be applied in defining a collectively financed basic health care services package.

In line with current usage in health economics, this report regards the CUA primarily as an efficiency analysis, with the QALY as a measure of health gain (CVZ99, Dru97, Hur00, Joh96, Ric94, Tor86, Uyl00). The report assumes that CUA is carried out from a general societal perspective, so that all effects and costs are relevant, regardless of who is involved or when they occur.

In scientific literature, the QALY is fairly generally regarded as the theoretically best grounded and, from a practical and methodological point of view, best developed outcome measure for CUAs. The report deals only briefly with possible alternatives, such as the DALY, the Healthy Year Equivalent (HYE) and the saved young life equivalent (SAVE). The numerous methodological choices that are relevant for the quality of evidence on the effectiveness of health care services *in general*, are not covered at all, although evidently they can have a major influence on the quality and usefulness of CUAs. Examples include the choices to be made in designing an effectiveness study, in the way it is carried out, and in the aggregation of the results of various studies in meta-analyses. The report also leaves aside the modelling of health care outcomes in the longer term: a field of study replete with challenges, pitfalls and traps, which is of vital importance for many types of health policy research, including efficiency analysis (Bar98, Bon91, Maa95, Man96a). In general, randomized clinical trials and other forms

of effectiveness research are of a few years' duration, at most. On the other hand, CUAs that are carried out from a general societal perspective try to consider the integral effects of health care services and often prompt statements to be made on their impact on the total course of life. Modelling of lifelong treatment effects, based on empirical effectiveness research with a limited duration, supplemented with other data, introduces significant uncertainties into CUA. This also applies to the forecasts for the related costs over the longer term.

The report is drawn up as follows. To start, Chapter 2 considers how the general guiding notion of efficiency in CUA is transformed into a quantifiable concept applicable across the entire health care field, indicating the ratio of the input ('costs') to the output ('health gain') of health care services. Chapter 3, the most comprehensive chapter, is devoted to the QALY: the outcome measure that distinguishes CUA from other forms of efficiency analysis. The route that must lead to a generally applicable measure of health gain appears to be paved with a series of assumptions on which the final word has not been spoken. Chapters 4 and 5 then briefly consider two components of CUA that are not specific to CUA, but that are inherent in every type of quantitative efficiency analysis. These are the quantification of the costs and the conversion of future health gain to its present value ('discounting'). Chapter 6 notes, equally briefly, what the effects of CUA-based allocation decisions are on the distribution of the achieved health gain. Finally, Chapter 7 explores the implications of the foregoing chapters as regards the applicability of the efficiency criterion.

Chapter 2

Efficiency

A cost-utility analysis represents a specific interpretation of the concept of efficiency. The economist Hurley noted in a recent health economic handbook that the term efficiency has a specific, technical meaning in economics that does not accord with its general use. He observed that, in daily usage, the word has a strong normative charge: it denotes "not wasting resources", something that one can scarcely object to (Hur00).

In health care, the notion that resources should not be wasted certainly has its own appeal and relevance. However, it is clear that this general, guiding thought lies at some distance from a tightly defined, let alone a quantifiable, efficiency concept applicable across the entire breadth of health care. Even the Health Council's definition for the concept of efficiency in Medical Treatment at a Crossroads (GR91) – "the benefits of an activity set against the costs in money, manpower, resources and time" – is still far removed from a similar general and quantitative concept. This also applies to the definition the Minister of Health, Welfare & Sport used in the policy note MTA and efficiency of care (Bor95): "efficient care is effective care, at the lowest possible cost". Neither of these descriptions entails the ambition of being able to compare the efficiency of types of health care that have different goals.

Obviously, there is a long route to travel to get from the normal meaning of the word 'efficiency', to its meaning in a CUA context. CUA's cost-utility ratio is a particular transformation of the informal efficiency concept. It arises from a quantitative model of the ratio of input to output of health care services and thereby represents a simplification of reality.

In this model, output is defined as health gain. More specifically, it is defined as the value the health gain represents for patients or potential patients. Various authors have referred to aspects of the value of health care that remain out of sight in this way (Ger93, Goo96, Moo94a, Moo94b, Moo00, Nee00, RGO01): the value of information *per se*, involved in the transfer of diagnostic, etiological, pathophysiological and prognostic insights; the value of reassurance; the value of respect, care and attention that the patients experience or, more generally, the so-called process utility of health care (Don97); the recovery or confirmation of their autonomy; the value of continuity in care (Shi97); the positive effects that flow from health care to those closest to the patients (Bar99, Bro01, Har01, Kor99, Loo89, Wei97a); the value that people attach to knowing that effective care services are available to everyone (Hur00, Moo00), and, more specifically, the value of preventing parental and societal concern by introducing vaccination against meningococcus in children; the value of recovery or improvement of someone's appearance (Bon01a); and, finally, the value of new human life (Bro93). These aspects are to some degree inherent in health care in general, as well as in the good treatment of patients and those closest to them. However, these are also, to a significant degree, ingredients and products of health care that has shown to be effective, that meets the requirements of the time and provides care that patients and care providers can trust. In short, these are, to an important extent, aspects where specific health care services may differ thoroughly from each other.

The input of health care is also defined in a specific way in CUAs. The basic assumption is that resources that are withdrawn from one form of care can be directly siphoned off to another, more efficient form of care (Dru97). The fact that the costs themselves also imply gains, in the form of the job satisfaction and income of care providers, is ignored. The definition of costs that may or may not be incorporated in the analysis (such as productivity losses or the costs of health care in extra years of life gained through medical treatment) is open to discussion (Bro01, Hur00, Koo98, Luc96). It is often passed over that the actual costs can be dependent on the scale on which a service is applied, on the experience that has been gained and on other time-dependent and place-dependent factors (Ash00, Ger93). As a rule, both the input and output of services are projected to a specific moment in time through discounting, without taking account of the dynamics of social and scientific development, which are unpredictable, but can be influenced to some extent by political and other decisions.

As we have noted, the C/U ratio stems from a model and is thus a simplification of reality. As is known, that is a model's strength and its weakness at one and the same time. A model offers an overview and insight that the complexity of reality would otherwise make difficult to acquire. It offers the possibility of mapping the considerations that are relevant to a specific decision. It may assist in bringing about an orderly discussion among the different parties involved in the decision process, by identifying points

of consensus and differences of opinion. In so doing, they may arrive step by step at an agreement that would otherwise have been difficult to achieve given the complexity of the problem and the many conflicting opinions and interests (Vle90). However, the risk is that the developers and users of the model may either consciously or unconsciously forget that they are working with a model. As a consequence, they may declare as out of order, or may not be able to weigh properly, any considerations that are either insufficiently addressed in the model or completely absent from it. Moreover, they may neglect the value laden choices that are involved in creating the model (Ber98).

Hurley, mentioned earlier, pointed out that technical economic concepts such as ‘efficiency’ and an ‘optimal’ distribution of resources in the public arena unavoidably acquire a normative meaning. Because of this, among other reasons, he argued that, despite any number of economists who would wish to see it as ‘objective science’, the so-called welfare economy (the economic tradition in which many economically oriented cost-utility analysts ground their work) is inevitably a form of social ethics and should be treated accordingly (Hur00). This is important to keep in mind in interpreting CUA-based statements about the efficiency of health care services. It implies that a cost-utility ratio is not to be taken as representing an objective state of affairs in reality, but instead as a value laden position in the debate about the distribution of scarce resources in health care.

Chapter 3

Utility

The word ‘utility’ in cost-utility can, but does not necessarily, indicate one of the meanings attributed to the term over time by philosophers and economists (Sen91). In daily health economic usage, the term ‘cost-utility analysis’ refers to efficiency analyses with a measure of health care output that meets two specific requirements. First, the output measure is applicable to the entire health care. Second, it represents the value attributed to the total of all the service’s health effects (Dru97). This chapter successively lists the various methodological steps involved in measuring the value of health gain in CUAs.

3.1 The ‘health gain’ concept

How is the output of a medical treatment to be quantified? The most common approach in estimating health gain in cost-utility analysis proceeds from the idea that the output of a health care service can be analyzed in two dimensions, duration of life and health-related quality of life. A medical treatment may prolong life, promote the quality of life, or both. Hence, it is possible to speak of a person’s ‘quality-adjusted duration of life’. This may be calculated by taking the duration of each period of life spent in a particular health state, multiplying it by the appropriate quality factor for this state (varying from 0 (dead) to 1 (entirely healthy)), and then aggregating these arithmetical products.* The outcome of a treatment for an individual, in comparison with no treatment, may then be

* To date, the fact that at least some people experience certain health states as being worse than death has not received much attention in the QALY methodology (Hou01, Mac01).

quantified as the difference between the quality-adjusted life duration with or without treatment.

This line of thinking lies at the heart of the QALY as a measure of health gain. A QALY represents one life year spent in a condition of perfect health, or two years spent in a health condition that is 'half as desirable', etc. A treatment that extends life by ten years while the health condition is 'worth half as much' as a condition of perfect health, and a treatment that does not extend life, but improves the health-related quality of life from 0.5 to 1 during ten years, both yield 5 QALYs.

Despite the attraction of its simplicity, the QALY concept has, right from the outset, led to critical questioning among cost-utility researchers and those less directly concerned. Is it reasonable to assume that one particular quality value can be attached to a specific condition of health, irrespective of the duration of the condition or of the situation that prevailed prior to it or that will succeed it, and irrespective of the possible alternative states that a (different) treatment might achieve? Is it reasonable to grade each healthy year of life equally, whether it is the final year of life or the first in a long series? Are QALYs of different people comparable to the extent that they may be tallied up when it comes to determining the outcome of a medical service for a given group of people? And if these propositions do not apply at all, or not in all circumstances, or not entirely, which consequences will it have for the interpretation of health gain measured in QALYs?

To date, the health economic literature and the related decision theoretical and ethical literature have consistently wrestled with these issues following the introduction of the term QALY in the 1970s (Ber73, p 196, Wei77, Zec76). Two global schools of thought, characterized by two different approaches to the QALY concept, are discernible among the health economic authors who have involved themselves in the matter. These are the welfare economic or 'welfarist' and the 'extra-welfarist' or 'decision maker's' approaches (Ble97a, Bro00a, Bus00, Cul98, Hur98, Joh96, Wei97b). The welfare economic approach strives towards the best possible embedding of the QALY concept in the theory of welfare economics, and thus attempts to create clarity in the interpretation of the QALY under various circumstances. In this vision, QALYs should be a valid reflection of the value attached to various health care outcomes by the individuals whose health is, ultimately, what health care is about. The decision maker's approach, in contrast, proceeds from the premise that those who must decide on the distribution of a limited health budget must conclude for themselves which 'objective function' they attempt to maximize with this allocation. As a rule, it is assumed that the decision maker's objective function will somehow represent 'health' itself and that at present the best conceivable representation of that maximand is 'the sum of the produced QALYs'.

Both approaches have their problems. The welfare economic approach charges itself with demonstrating that the QALY, under specific conditions, can be used as a measure of health-related 'utility'. The idea is that people strive to maximize their individual preference function or 'utility', that the overall welfare of society is a function of these individual utilities, and that the aim of the distribution of scarce resources should be to maximize the social welfare function (Gar96). The concept 'utility', to which conflicting meanings are attributed in the philosophic and economic literature (Sen91), is generally used in the welfare economic literature in the sense of the expected utility theory. The foundations for this theory were laid in 1944 by Von Neumann and Morgenstern (Neu44). The standard gamble method* derived from this theory prevailed for a long time among health economists as the most theoretically pure way of measuring individual health-related utility. In recent years, some health economic researchers have disputed the importance of the expected utility theory as a theoretical basis for the QALY. They acknowledge its value as a *normative* theory. However, they contend that its known *descriptive* shortcomings (brought to the fore in the work of authors such as Kahneman and Tversky (Kah79)), disqualifies the expected utility theory as a theoretical foundation for the descriptive exercise that they consider the derivation of preferences for health states from people's choice behaviour to be. They believe a nonexpected utility theory, such as the rank dependent utility theory, is more in agreement with people's preferences for health states. This theory would therefore offer a sounder basis for developing instruments to measure those preferences (Ble97b, Ble02).

In 1980 Pliskin and colleagues formulated three conditions that had to be fulfilled if the QALY was to be interpreted as a unit of utility (Pli80, Wei80):**

- the valuation of the quality of life and the duration of life must be independent of each other (utility independence)
- the proportion of the remaining lifespan that one is prepared to sacrifice in the interest of a specific quality improvement must be independent of the length of the remaining lifespan (constant proportional trade-off)

---

* See the first footnote with paragraph 3.3.3.

** At the World Congress on Health Economic in the same year, Alan Williams listed seven assumptions that were, to a certain degree, comparable: "very stringent conditions, which make empirical work to elicit such valuations rather difficult. They are needed in order to avoid the logistically impossible alternative, which is to present subjects with all possible future time profiles of health states, and then to get them to value each profile as a whole relative to each other profile as a whole." Mehrez and Gafni later propagated Williams' 'logistically impossible alternative' in a somewhat modified form in the Healthy Year Equivalent described later in this paragraph.

- there must be 'risk neutrality on life years'.* It is assumed, for example, that one is indifferent between the certainty of nine healthy years of life on the one hand, and participation in a lottery with a 90% chance of ten healthy life years and a 10% chance of sudden death on the other hand.**

Subsequently it was also noted that in the situation of a health condition that varies over time (which often occurs in practice), the value attributed to the health state in a specific period must be independent of the health states in all other periods. This condition is known as additive utility independence or additive separability (Ble95, Joh96, Dol00). Dolan also distinguished the requirements that the value attached to health states need to be stable during someone's entire life (stability of lifetime preferences) and independent of the moment (close at hand or far in the future) at which they occur (zero rate of time preference) (Dol00).

These are not minor requirements and diverse authors have argued that they will generally not be met in practice (Ble95, Boe98, Bro93, Car89, Dru97, Gaf95, Gar97, Goo96, Gui99, Hey90, Loo89, Nor94, Tor89). Empirical research also provided many indications pointing to this. The evaluation of a health state appeared to be dependent on a number of factors: the length of time spent in the state (Kir00, Sac78), the prospects of improvement in, or deterioration of, the condition (Kra98a, Ric96), the long-term prognosis, and the cause of death (Hal92). Dolan offered an overview of a large number of studies that investigated the tenability of one or more of the six named requirements. He concluded that no clear empirical support could be found for any of them (Dol00). In particular, the assumption of additive separability appeared to be contradicted persuasively by the available information. This is in conformity with Broome's argument, which qualified precisely this requirement (amid various "heroic assumptions") as "the most dubious condition" (Bro93). The provisional conclusion of the welfare economic discussion on the QALY seems to be adequately summarized in the book by Drummond and colleagues. They answer the question "Is a QALY a utility?" with a concise "In general, it is not" (Dru97).

To address this fundamental problem, Mehrez and Gafni introduced the Healthy Year Equivalent (HYE) in 1989 (Meh89). The QALY is based on an assessment of

---

* Bleichrodt and colleagues later showed that the third requirement, that of the risk neutrality, *implied* the other two requirements under the (entirely plausible) requirement that all health states are equal for a duration of zero life years. They emphasized that this conclusion should not be taken as a *defence* of the QALY concept. Indeed, in terms of the plausibility of the three conditions together, it seems to make little difference whether utility independence and constant proportional trade-off can now be considered as independent requirements, or whether they are included in the assumption of risk neutrality.

** It is theoretically possible to adjust the calculation of QALYs to non-neutral risk attitudes, but this requires an (awkward) estimation of the relevant risk-attitude parameter (Kra98b, Miy85). It is rarely done in practice (Joh96).

health states without specified duration. In contrast, respondents that are measuring the health-related utility in HYEs are asked to determine the degree of desirability of a total lifetime health profile (that is to say, of a specific sequence of health states, each of a specific duration, through to death). The determination of utility in terms of HYEs cannot occur either without assuming risk neutrality with regard to a (healthy) lifespan (Dol00, Rie98). For the rest, however, the assumptions underlying the HYE are much less restrictive than those of the QALY. The theoretical superiority of appraising lifetime health profiles instead of isolated health states has also been fairly generally recognized in the literature (Ble95, Bon92, Bra99, Bus00, Dol00, Dru97, Gol96a, Rie98). However, the conclusion that the practical feasibility of the HYE approach is dubious has been widely shared as well. The appraisal of one sequence of various health states is already difficult, but the appraisal of an outcome scenario embracing a large number of different lifetime health profiles is an immense task. It obliges extended interviews and places high cognitive demands upon the respondents, with all the consequent problems in terms of costs, recruitment of respondents, and reliability and validity of the outcomes produced (Bon92, Dru97, Joh96, Ric96). Therefore, the HYE approach has been seldom applied in empirical research (Lle02).

Some welfare economic oriented health economists has suggested that the assumptions underlying the view of the QALY as a unit of utility do not perhaps entirely hold up in practice, but do so approximately and sufficiently well to justify using QALYs in CUAs (Gar96). However, to date, the (partial) empirical testing of this claim, which is theoretically possible through a comparison of the QALY and the HYE approach, has not been forthcoming. There are hardly any studies available about the degree to which the results of the more current quantification of health gain in QALYs diverge in practice from those of the theoretically superior quantifying in HYEs (Bra99).

Thus, the interpretation of the QALY as a unit of individual utility encounters serious problems. In addition, the aggregation of individual health-related utilities to a value judgement at the societal level, which forms the keystone of the welfare economic approach to the QALY, is not unproblematic either. Whether, and under what conditions, it is possible to aggregate individual utilities in a societal utility function is the core of an intensive debate in welfare economics that also reverberates in health economics (Fro89c, Gar96, Sas01, Tor86, Wag91). Some authors suggest that the aggregation of QALYs is possible under terms that are not too limiting (Ble97a), while others believe that the necessary interpersonal comparability of QALYs cannot be sufficiently guaranteed (Bro00a, Ube00c). The aggregation of health state valuations is further complicated by the fact that an appreciable number of people assess the condition ‘death’ higher than some health states (Bon01a, Mac01). The debate continues.

Given the stubborn theoretical and empirical problems with which one wrestles in the QALY’s welfare economic approach, it is hardly surprising that the extra-welfarist or

decision maker's approach has enjoyed growing popularity in recent years (Bus00). The latter's virtue is that it does not seek an underpinning in individual utilities, and – at least in the eyes of some authors – allows a more pragmatic common sense approach (Bro00a). The extra-welfarist view assumes that decision makers strive for a 'maximization of health' in distributing health care resources, and that 'maximizing the sum of the produced QALYs' is an acceptable interpretation of this goal (Bus00, Dru97, Wag91). At the same time, adherents of this approach are aware that 'maximizing QALYs' may well be only one of the decision makers' goals. Other considerations, such as distributive justice, degree of dependence and the right to privacy, may play a role as well (Bus00, Rus96, Wei97b). The adherents to this approach emphasize, and in so doing claim to give expression to "the prevailing view of cost-effectiveness analysis", that CUA must be seen as "an aid to decision making, not a complete prescription for social choice" (Rus96, Wei97b).

Is the extra-welfarist view of the QALY an alternative to the welfare economic view? Why should the QALY succeed as a measure of health while it fails as a measure of utility? One only needs to ask what is meant by multiplying lifespan by health to realize that a critical questioning of the QALY as a standard for measuring the 'amount of health' inevitably leads back to all the complexity that encompasses the concept of the QALY as utility. It may also hardly be surprising that it does not appear to make any difference in the practice of cost-utility research whether one quantifies welfare economic or extra-welfarist QALYs. Hurley found that "[...] there is nothing observable about the methods themselves to indicate whether the analysis is intended to be welfarist of extra-welfarist. Nothing observable distinguishes which approach is being invoked" (Hur00). In practice, all researchers appear to operate as if their interpretation of the QALY is a welfare economic one. They attempt to measure as well as possible the value that people from the general population attach to health states and then use the resultant values as factors in weighting the lifespan. A large part of the methodological research is aimed at developing techniques for adequately quantifying the individual preferences. Implicitly, the majority of cost-utility researchers appears to support the welfare economic interpretation of the QALY (Ble96). Nevertheless, the same majority, in accordance with "the prevailing view of cost-effectiveness analysis", puts the responsibility for *weighing* the thus-interpreted QALYs in the decision making process into the hands of the decision makers.

And what should these decision makers make of this? Hurley remarked that, ironically, the attempts of welfarist and extra-welfarist oriented cost-utility researchers to provide clarity on the assumptions and meaning of QALYs have only succeeded in increasing the confusion among the potential users of the CUA outcomes. He concluded, paraphrasing a well-known health economic adage (Wei88, Wil92): "It turns out that a QALY is *not* just a QALY". Sometimes the QALY appears to be a measure of

'utility', at others it is a measure of 'health'. "The interpretation is in the eyes of the beholder, depending on what assumptions one is willing to make" (Hur00).

Meanwhile, cost-utility researchers appear to let themselves be governed by the conviction, so eloquently articulated by Williams, that there is simply no better available alternative to the QALY:

> "[...] there is no perfect system on offer, and we can't wait. As with a well conducted clinical trial, the new has to be compared systematically, according to preselected criteria, with what already exists. This is what needs to happen in the field of priority setting. If the same criteria as are used to criticise the QALY approach were used *in an even handed way (italics in original)* to criticize current practice, or any feasible alternative to it, how would these alternatives make out?" (Wil96, p 1801)

Interestingly enough, this quotation does not only show how this prominent economist regards the current practice of setting priorities in healthcare. It also reveals how he interprets the QALY approach, despite his extra-welfarist arguments (Sug78, Wil93): indeed as an actual alternative to the current practice of priority setting, rather than as a modest supporter of it.

The decision makers have the ultimate say. They will primarily need to go by their own insight and intuition. Empirical research that could shed light on the importance of QALY assumption violations in various circumstances is either scarce or completely unavailable (Bra99). They will need to ask themselves a few questions, one of them being whether it is sensible to use one outcome measure to capture both differences in health state value and the difference between life and death. In a critical analysis of the DALY, which is also largely applicable to the QALY, Anand and Hanson wrote:

> "While having an indicator that combines states of imperfect health with death is clearly convenient, there is an obvious information loss in reducing death to simply another health state. Some will argue that the events are incommensurable, and that a lexical priority attaches to life over death. At any rate, this suggests that information about mortality and morbidity should be presented separately – even if trade-offs were conceded between the two events." (Ana97)

3.2 Describing health states

In order to measure the output of a medical treatment in QALYs, the health-related quality of life needs to be quantified. This generally occurs in two phases. First, patient-related research is carried out that leads to a *description* of the health states of those who have undergone the treatment in question. Second, the descriptions of the health states that are at issue in the treatment are *valued* on an interval scale* from 0 to 1. In principle, it is possible to determine the quality adjustment factor necessary for calculating QALYs in the patient-related research directly, i.e.: without the interim step of the health states description. In that case, the patients involved in the research are asked to articulate a valuation of their own health state(s). This is termed a direct utility assessment. However, this approach is used extremely rarely in practice. This is because most researchers assume that it is not the actual patient but (a random sample of) the general public that constitutes the most appropriate source of health state valuations in CUAs (see 3.3.2) and also because clinicians often regard a direct utility assessment as an unacceptable burden on patients (Bra99).

Since the late 1940s, an intense discussion has been going on in the medical-scientific literature, on how to study the effects of medical treatments on patients' health states (Fei87, Kar48, Kar49, Nee91). Especially in the past 20 years, many hundreds of instruments have been developed for measuring 'health states', 'health-related quality of life' or aspects thereof (Spi90, Spi96). The measurement instruments mostly take the form of a list of questions about aspects of 'health' or 'quality of life'. The questionnaire may be filled in by the patients themselves or by interviewers.

Broadly speaking, four types of measurement instrument can be distinguished. First of all, there are the domain-specific instruments (Ess00). These instruments cover one domain of health and can be applied to various diseases. Examples are the index of independence in Activities of Daily Living (ADL index) (Kat63) and the McGill Pain Questionnaire (Mel75).

Then there are the disease-specific instruments. This was the most commonly used type of instrument in a review of 154 health-economic evaluation studies (Bra99). These instruments are generally developed for medical research to evaluate treatments for one specific disease. They focus on the dimensions of the concept of health that are affected by the disease at issue. Examples are the Health Assessment Questionnaire developed for research into people with rheumatic conditions (Fri80), and the EORTC QLQ-C30

* An interval scale is a scale constructed in such a way that a quantitative meaning can be attributed to the intervals between the values on the scale (for example, "The difference between A and B is twice as large as the difference between A and C"). However, because no absolute zero is defined, the ratios of the values themselves are meaningless ("A is twice as great as B" is a meaningless statement).

(the core questionnaire of the European Organization for Research and Treatment of Cancer), which can be complemented with separate modules for specific forms of cancer (Aar00).

The third category encompasses the instruments aimed at representing the health concept in general, the general health state instruments. These psychometric-style instruments are also frequently used in economic evaluation research (Bra99). They consist of a large number of questions ('items'), each representing a particular aspect of the complex concept of health. The scores in the individual items can be added together to total the most important 'dimensions' of health (for example the physical, the psychological and the social dimensions). These dimensions may in turn, at least with some instruments, be totalled to achieve a final score for health as a whole. In the summation, the item scores and dimension scores may be weighted differently. Various methods are used to determine the weighting factor. It has been based, for example, on empirical research into the statistical correlation between the scores of related items in the SF-36 (War96). In the Sickness Impact Profile, it has been determined by asking a panel of assessors to rate each item on a 15-point scale from minimally to maximally 'dysfunctional' (Ber76). Similarly, a panel has assessed the 'seriousness' of the items in the Nottingham Health Profile (Hun80). Although a form of weighting separate item scores is thus unavoidably at issue when presenting the measurement results, the general health state instruments are usually regarded as descriptive instruments (Ess95).

General health state instruments are distinguished in this respect from the fourth and last category of instruments, the health state classification systems (preference classification systems (Gol96a), multi-attribute health status classification systems (Dru97), multi-attribute utility scales (Bra99)). These instruments embody both the descriptive and the valuation phase of the research necessary for rating the quality adjustment factor of the QALY. The valuation phase will be covered in the next paragraph. It is sufficient here to note that most health state classification systems, in contrast with the general health state instruments, are based on a classification of health into a relatively small number of discreet states. Examples are the Health Utilities Index, which in its third and most extensive version distinguishes 972,000 different states (Fee96), and the EuroQuol (Eur90), which contains 243 states in its EQ-5D version (Kin96). By comparison, the SF-36, one of the most used general health state instruments, describes more than 10 million possible health states (Dol00). Although health state classification systems have been available for some 25 years, a glance at, among others, a review of CUAs published up until 1997 shows that they are relatively rarely used in economic evaluation research (Bra99, Neu97, Neu00a).

There is broad agreement in the literature on the advantages and disadvantages of various types of measurement instruments when it comes to their application in CUAs. The disease-specific and domain-specific questionnaires are sensitive instruments for

measuring the symptoms for which they are designed (Ess00, Gol96a, Guy96, Pat89). An obvious disadvantage for economic evaluation research is that they do not allow any comparison of the effects of treatment for diseases that influence different dimensions of health. By contrast, the general health state instruments enable a description of health effects in terms that are applicable to many different diseases, albeit at a cost of reduced sensitivity to some subtle or disease-specific effects and changes (Ber89, Edg98, Guy96, Pat89, Ver01). For both the disease-specific and domain-specific, as well as the general health state instruments, it is still unclear how to achieve the valuation of the resulting health state descriptions that is necessary to calculate QALYs (Bra99, Dru97). The health state classification systems have been specifically developed to offer a solution to this latter problem. However, they fall short even more than the general health state instruments when it comes to detecting subtle and disease-specific symptoms and the changes therein (Bra99, Bus00, Don88, Gol96a, Guy96, Hol95, McK98, Neu00a, Tor99, Ver01).

All authors agree that the ideal type of instrument that could be broadly employed across all disease categories to guarantee the *comparability* of the health gain measured is not yet available. Various solutions have been proposed. The health state classification systems form a part of all of these, but the disease-specific and general health state instruments still seem to have a role as well. A lack of clarity remains in particular on the issue of how the results of these various instruments should be combined in one dimension.

The authors of the Canadian guidelines for pharmaco-economic research (CCO97) (taken as a starting point in the guideline development by the Dutch Health Care Insurance Board (CVZ99)), suggested that it is unlikely that one instrument will ever achieve 'golden standard' status for measuring the health-related quality of life. They recommended that pharmaco-economic researchers select one instrument from each of the four categories (the disease-specific and domain-specific, the general health state instruments and the classification systems), taking into account the study-specific considerations at issue in such a choice. They leave aside the question of how the results of these measurements can be converted into weighting factors for QALYs that are comparable over various studies (CCO97).

Rotterdam researchers recommended utilizing one of the health state classification systems for calculating QALYs, but they also insisted on incorporating a specific and a general health state instrument in economic evaluation research. Just as with the Canadian guidelines, they did not address the question if and, if yes, how the obtained measurement results could be involved in the calculation of QALYs that are comparable across various studies (Bus00, Ess00).

The American Panel on Cost-Effectiveness in Health and Medicine, the so-called Washington Panel, concluded that "the field of measurement and valuation of outcomes

is still a developing one". It advocated the development of one generally applicable health state classification system that is better than the existing systems. At the same time, the Panel stated that even an ideal instrument would probably not render the disease-specific instruments redundant. It described in general terms how the results of a small subcategory of the disease-specific instruments, the disease-specific classification systems, could be projected onto the dimensions of a general classification system to enable the calculation of QALYs (Gol96a).

The compilers of an extensive literature review that appeared three years after the publication by the Washington Panel showed that the methods used to value the results of disease-specific and general health state instruments in an economically valid way are still in their infancy. They expressed little optimism on whether they will ever be workable in practice (Bra99). The author of the chapter on measuring the health-related quality of life in a health economic handbook was very explicit in his conclusion: neither the disease-specific nor the general health state instruments are suitable for use in CUAs. When it comes to comparing outcomes for divergent diseases and treatments, then only the health state classification systems are usable (Dol00).

It is clear that the last word has yet to be spoken on this subject. For the time being, discussions on which instruments to use in CUAs for measuring health states are in an awkward predicament. On the one hand, there is broad agreement on the fact that disease-specific instruments are essential in order to avoid overlooking relevant treatment effects and to value them properly. On the other hand, major importance is attached to the generally applicable (but for specific effects, relatively insensitive) health state classification systems, because use of those systems seems to be the only workable way to generate health state values that are broadly comparable.

In the meantime, the practice of CUAs appeared to have taken little notice of the growing agreement on the importance of the health state classification systems. Most of the studies published between 1975 and 1995 made use of disease-specific instruments and derived the values that are attributed to health states from the researchers' judgement (Neu97). The health state classification systems also appeared to be only rarely utilized in CUAs that were published up to 1998 (Neu00a).

How important is the discussion on the choice of a measurement instrument for the decision maker considering (partly) basing allocation decisions on CUAs? The decision maker can derive little support from the literature for determining to what extent efficiency estimates are influenced by the choice between a general classification system and an disease-specific instrument. One study into the cost-utility of mammographic screening for breast cancer indicated that the choice *could* be of enormous practical importance. In that study, which used disease-specific descriptions of health states, the cost per quality-adjusted life-year gained appeared to be more than twice as high as the cost for the unadjusted life-year gained. The researchers ascribed the failure to find a

similar difference in another study to the fact that, in that study, a measurement instrument was used that was insensitive to the specific harmful health effects associated with the diagnosis and treatment of breast cancer (such as anxiety about the diagnosis, pain, stiffness and swelling that could be the result of breast amputation, fear for recurrence of the cancer and worries over appearance) (Hal92).

Neither in choosing between the available classification systems is the existing literature of much help. The authors of an earlier-mentioned literature review pointed to the clear differences in descriptive content between the individual classification systems and suggested that the selection of a particular system should be made dependent on the disease and age of patients. Comparative research into the measurement results of the various classification systems among the same patient populations is still extremely rare and difficult to interpret (Bra99). Thus, two research groups that compared the Health Utilities Index and the EuroQol instrument found *differing* agreements and differences (Bos00, Gli99).

The complexity to which the decision maker is subjected is further increased by the problem of the intercultural comparability of the measurement results. Almost all questionnaires for measuring health states are developed on the basis of research that almost exclusively involves respondents who are from the relatively well educated, white majority of a western nation. It seems plausible that the questionnaires, with their specific selection of health aspects and their specific weighting of the item scores, will not necessarily yield comparable results if applied in other (sub)cultures, national and international. Research has confirmed this (Hut96, Hun98a, Ste00). In different cultures, there are different expectations of what it means to feel good, or to be healthy, or to be ill. These expectations also form and interact with perceptions of pain, discomfort and limitations, and the relative seriousness of symptoms. Furthermore, there are differences in the type of language which is used when recounting personal experiences and in the conventions governing the communication of those feelings to other people (Hun98a). Organizers of major international clinical trials have for years been pointing towards problems with the intercultural (non-)comparability of health state measurements (Spi96). This has, however, still received little attention in the discussion on CUA methodology. For example, the Washington Panel stated that one population group's articulated *preferences* for health states are not necessarily valid for other, differently constituted groups, but it ignores the comparable problem involved in *measuring* health (Gol96a). This problem is not mentioned in Dolan's literature review either (Dol00).

3.3 Valuing the quality of life

3.3.1 Nature of the valuations

Do people entertain preferences regarding health states and if so, what is the nature of such preferences? This is a question that is only rarely posed in the health economic literature. Dolan's argument is an instructive exception (Dol00). Following the psychologist Fischoff, Dolan distinguishes between the philosophy of articulated values and the philosophy of basic values (Fis91). Health economists proceed from the premise that the preferences regarding health states that they measure among their respondents are also the 'true' preferences of these people. Their assumption is that the respondents have well-defined preference functions that can be 'tapped into' by appropriate questions.

This philosophy of articulated values is at odds with the results of numerous studies that indicate that subtle changes in the elicitation procedure appear to change the stated preferences of respondents. Such changes could include, for example, changes in the structure of the choice problem being put to the respondent, the question format and the sequence in which good and bad health states are presented.

This sort of findings supports the philosophy of basic values. This states that preferences in everyday life, as well as in health-economic research, "are not simply read off some master list, but are constructed on the spot by an adaptive decision maker" (Slo95). In this view, the elicitation procedure is a major force in shaping the preferences that the respondents express. Apart from the measurement procedure, respondents entertain only unarticulated, 'basic' values with regard to health states.

A middle way is conceivable between the philosophy of articulated values and the philosophy of basic values. Instead of just having preconceived, directly accessible preferences regarding health states, people could have 'stable values of moderate complexity' (Fis91). These values could be given further form and brought to light in an elicitation procedure that calls on the respondent's reflexive capacities. Dolan declares himself in favour of such a middle way, which makes it possible to proceed from the premise that the measured health state preferences do reflect the 'true' preferences of the respondent in question, but also induces the development of more complex measuring procedures. These will in most instances be more interactive procedures than are the norm now. Instead of simply 'tapping into' already-present preferences, these procedures will address the fact that the respondents' preferences must embody their final form during the measurement (Shi97). This middle way enables a wide spectrum of approaches. But, Dolan emphasized, given that most people are only familiar with a few different health states through their own experience, and that they have no well-defined preferences regarding many states, future studies would do well to select a starting point

closer to the philosophy of basic values than to the philosophy of articulated values (Dol00).

Against the background of Dolan's discussion of the influence of the measurement procedure on measured preferences, it is notable how little weight is attached in the health economic literature to the unrealistic character of the choice situations in which respondents are expected to imagine themselves when using the most important health-economic procedures to generate health state valuations* (Joh96). Social scientific and medical literature points to the abstract and artificial character of the choices that the respondents are confronted with. It is questioned that the quantitatively measured preferences can make any claim to being valid reflections of the unarticulated, changeable and context-dependent ideas about the value of health states that people actually entertain in reality (Ash89, Car89, Lep97, Nee92). Rather, critics argue, the measured values of health states are artefacts. If they are to be *relevant* artefacts that can credibly fulfil the role that they are assigned in the calculation of QALYs, then at the very least the questions posed to the respondents must make clear for which goal their answers are to be used. "Without some such explicit link between the questions used to establish quality indices and the allocations generated by QALYs, QALYs will remain persistently suspicious"(Men90).**

This criticism has received little attention in the health economic literature. Someone who was actually receptive to it is the Norwegian researcher Nord. He launched the 'saved young life equivalent' (SAVE) as an alternative to the QALY in 1992 (Nor92a). His proposal was received with appreciation in the Netherlands (Leu93) and was often cited in recent health economic reviews as promising but not yet adequately worked out and insufficiently tested in practice (Bus00, CCO97, Dru97, Gol96a). Nord's idea was partly inspired by the observation that the measurement procedures used to quantify the Q of the QALY in no way reflect the nature of the choices that make up CUA's reason for existence (Nor92a, Nor94). He postulated the SAVE as a new unit to measure the output of health care. A SAVE is produced if one saves a young person from a certain death and restores him or her to full health. He suggested asking respondents in health economic research how many outcomes X of a given medical intervention are equivalent to one SAVE. In this calculation, X represents the health gain for one patient, defined in terms of the health state prior to the intervention, the improvement of the health state, the risks of the intervention, the age of the patient and other relevant variables. If the answer is ten, then the value of outcome X is equal to one-tenth of a SAVE.

---

* For a description of these procedures see the first footnote of paragraph 3.3.3.

** Concerning a comparable problem (the weighting of health states by experts calculating DALYs) Anand and Hanson stated that "[...] the meaning attached to the different weighting of health states depends in an important way on the precise question that was asked of these experts. Their responses would also depend on understanding the use to which such estimates would be put"(Ana97).

Nord's SAVE procedure explicitly conceives the health economic valuation problem as one of the relative valuation of the health benefits that people experience from divergent health care services (Nor93, Nor94, Nor95, Pin97). A concomitant advantage is that the SAVE is based on a direct assessment of specific *changes* in the health, rather than, as in the QALY, on assessments of health states from which the value of the transition from one state to another is then derived (Nor99, Ube00a). In addition, the SAVE escapes the limiting assumptions discussed in 3.1 that lie at the heart of the QALY. Moreover, it provides, in principle, a solution to a number of distributional problems associated with the cost/QALY ratio (see Chapter 6), since the question as to how to value health benefits accruing to the young and to the old, to the seriously ill and to the almost-healthy, and to a few individuals and to many, is part of the valuation problem put before the respondent.

These advantages do come at a price at the practical level. The assessment task that confronts the respondents in the SAVE procedure is even more complex than the task that must be fulfilled to estimate Healthy Year Equivalents (see 3.1). Nevertheless, the advantages in comparison with the QALY are of such a nature that it is regrettable that so far the SAVE has been scarcely used in the field of health economic research, apart from some theoretical and methodological contributions to the development of the person trade-off approach that forms the procedural core of the SAVE proposal (Gre01, Nor99, Ube99, Ube00b, Ube00c). When it comes to quantifying the efficiency of services across the entire breadth of health care in one dimension, the development of complex, interactive procedures in which a relevant panel gradually forms a judgment on the value of the output of those services in terms of SAVEs, seems to be an approach that deserves thorough exploration. Such a way of proceeding would be in accordance with both Dolan's argument and the call by Edgar and colleagues for a discursive-ethical approach to the valuation problem (Edg98).

In contrast to the SAVEs, the QALYs, as they now figure in cost-utility research, are based on abstract valuation exercises that bear no relation to the use to which the elicited valuations will be put. This casts a shadow over the applicability of the outcomes of this research.

3.3.2 Choosing the valuators

Whereas little has been written on the nature of health state valuations, there has been much discussion on the question as to who they must be elicited from. The two most discussed options are the general population and the patients who are now undergoing or have undergone the disease and the treatment in question. Other possible sources of health state valuations have been named incidentally, but were generally quickly disposed of. This applies to the researchers who are carrying out CUA. In the eyes of

authoritative authors, they form a source of quality adjustment factors that can barely be taken seriously (Joh96). Nevertheless, it was the most frequent one in a series of CUAs published between 1975 and 1995 (Neu97). It also applies to health care professionals. In an era in which doctors are primarily valued not so much as moral authorities but rather as the bearers of medical-scientific knowledge, the idea that they should express value judgements about the quality of life of their patients can expect little approval.* Doctors, nursing staff and other care providers certainly have insight in and experience with various diseases and treatments, but research shows that their judgement on the (un)desirability of a state is often at odds with that of their patients (Fro89c). In addition, and for several conscious or unconscious reasons, they can incline towards an overestimation of the seriousness of disease states in which they themselves are professionally interested (Tor86, McK98).

This leaves the patients and the general public. Dolan correctly indicates that these are not mutually exclusive categories. There are many people among the 'healthy' public who know several diseases through their own experience. In addition, there are patients who have only just become ill or who were instead ill a long time previously (Dol99, Dol00). What it ultimately comes down to is a decision on how the preferences of different population groups must be weighted (Loo89) in cases where systematic differences in valuing health states become apparent between the groups. Are there such differences? Some studies show that people who are themselves in a bad state of health value their own condition more highly than others do (Boy90, Len99, Sac78). No differences are found in other studies. A 1989 review article determines that the available research results does not admit a clear conclusion, but that genuine differences could very easily have gone unnoticed because of the limited scope of most studies and the large interindividual variability (Fro89c). A clear difference is indeed later found in a relatively large study where patients valued varying health states higher than non-patients did (Dol96). A recent review of 38 studies confirmes that observation (Wit00). In fact, most authors proceed from the premise that one must take serious account of the possibility that patients rate their own condition more highly than people who have no experience of it do. This could particularly be the case among people with a chronic disease. This difference could be explained by healthy peoples' unfamiliarity with disease

* Murray and colleagues, in opposition to the prevailing views in the health economic and ethical literature about CUAs, used panels of health care professionals to determine the disability weights for calculating disability adjusted life years. Murray defended this on practical grounds: "To speed up the already demanding process, the protocol [for weighting health states] has been designed to be used with health care providers so that less time needs to be invested to describe each of the conditions. Health care providers are selected because of their knowledge, not because they have 'better' judgement" (Mur97). The Dutch Disability Weights group also appealed to medical experts, particularly general practitioners (Sto00a).

states, but certainly also by successful adaptation of patients to their condition (Dol00, Gol96a, McK98, Men99).

Which valuations should weigh more heavily, those of the patients or those of the general population? No one disputes that only the patients themselves know what it is to be in a particular health state. They have access to direct insight, which is denied to others, into the feelings and limitations that are associated with their condition. Compared to patients, non-patients have an immense 'information disadvantage' that can not be bridged with even the must subtle description. But no matter how important this is, certainly from a practical viewpoint (Tor86), the crucial main question is still whether it is the *values* of patients, or those of the general population, that are relevant to the decision making on resource allocation which is central to CUA.

For the patients, or more generally the health care service target groups, the argument is that they are the ones whose welfare is what health care is all about. It thus hardly seems defensible to ignore the patients' values. According to some health economists, this line of reasoning accords best with the principles of welfare economics, which state that allocation decisions must be based on the utility functions of those who could benefit from the decisions (Bra99, Joh96).

The possibility that patients could be inclined towards strategic response behaviour argues against this approach. Patients could be tempted to overestimate the undesirability of their condition and in this way accord a higher value to the efficiency of services aimed at the prevention or curing of their condition (Loo89, Tor86). Others doubt whether respondents are in a position to work out in practice how their answers could influence the allocation of resources. Furthermore, available empirical research suggests that patients are quicker to overvalue, rather than undervalue, their own health state (Joh96, Gol96a). Even more important is the objection that patients, to the degree that they have successfully adapted to their condition, will express their adapted assessment in their responses. The question is whether it is exactly these *adapted* values that the valuation of health states in CUAs should be based upon (Bro95, Dol00, Men02).

Those who argue in favour of the general population as a source of health state valuations answer this last question negatively. They state that the allocation of collective resources in health care is the affair of the entire society, in which patients with a particular disease form only a small minority. The allocation decisions must express the values of the large majority of healthy people who are entirely able and competent to determine (without knowing which diseases they will suffer in the future) into just which services their insurance premiums should be invested (Bus00, CCO97, Gol96a, Rut98). The fact that they are only inadequately aware of the actual nature of health states does not detract from the validity of their value judgements. After all, incomplete and imperfect information is an inescapable fact of life (Loo89).

It can be noted against this argument that even though it may be entirely legitimate that health states are assessed by representatives of the general population, this still does not mean that their valuations that are the *right* ones from an ethical point of view. Dolan posed the question of whether it is defensible to attach equal value to the judgement of someone who has never given any thought to a particular disease as to that of someone who has direct personal experience of it (Dol99). Healthy people without the experience of particular diseases or limitations may be guided by stereotypes or otherwise misplaced prejudices (Gol96a). Froberg and Kane provided a number of examples of research results that showed the values that the respondents accord to health states to depend on the question as to whether the states concern themselves or hypothetical patients. In the case of a hypothetical patient, they would appear to be dependent on various patient characteristics that do not actually have anything to do with the health state itself. The authors concluded that "the results of these studies highlight the serious ethical considerations that arise when social preferences are used to make public policy decisions" (Fro89c).

Another counterargument is that it would be unwise not to take account of the way in which people generally value their condition once a particular disease or limitation has manifested itself in them. This is also the case when the logic that the insured party certainly has the right to determine what he or she is insured for, is considereed as being applicable without reservation to the distribution of collective health care resources (Bro95, Men02). For people want to insure themselves for health care which best meets their needs once they need it.

The authors of guidelines for health economic evaluation research, such as those of the Washington Panel, the Canadian and Dutch guidelines for pharmaco-economic research and the guidelines of the Rotterdam institute for Medical Technology Assessment, have all adopted a clear standpoint in this discussion. They have stated plainly that health state valuations must derive from random samples of the general population (Bus00, CCO97, CVZ99, Gol96a). This guideline is also put into practice in developing the health state classification systems discussed in 2.3 (Bra99). However, the writers of recent literature reviews who do not pose themselves the task of formulating guidelines, originating from both ethics (Edg98, McK98) and health economics (Bra99, Dol00, Joh96), do not regard the question of the choice of the relevant valuators as in any way settled. The continuing debate in recent literature would appear to justify their position.

Nord and colleagues have made an interesting proposal. They suggest distinguishing between the utility of health states (which the patients can best judge themselves) and the social value of changes in health states brought about by medical services (which can best be determined by a random sample from the general population). According to these authors, the utility that the patients themselves ascribe to their own health states should be weighed by the representatives of the general population in their assessment

of the social value of health gains (Men99, Nor99, Ube00b, Ube00c, Men02). To determine the social value, this proposal calls on the person trade-off approach, which has up to now enjoyed little popularity among health economists. Thus, it cannot be expected to find much of an audience in the short term. However, it certainly does cast a clarifying light onto the discussion about the choice of valuators, even if it does not offer a solution to the substantial ethical question as to what weight should be attached to patients' utility estimates influenced by successful adaptation (Men02).

To what degree does this discussion have practical importance? Can the extent to which the valuations of general population and patients are taken into account have consequences for the ranking of services according to cost utility? A study into the assessment of the quality of life in interventions to prevent osteoporosis showes that the influence might be considerable. There appeared to be important differences in the valuation of osteoporosis-related health states by women who have experienced osteoporotic fractures compared with women who have not. The estimated cost-utility of osteoporosis prevention thereby varied from $25,000 to $100,000 per QALY gained, depending on whether the values were used of the women without or with fractures (Gab99).

Just as in the measurement of health states, it seems that cultural diversity may also be an important factor in choosing the valuators of health states. Little attention has been devoted to this in the literature (Bad01, Gol96a).

3.3.3 *Choosing a measurement procedure*

To an even greater extent than the choice of the valuators, the choice of the right procedure for measuring health states preferences is a methodological subject in cost-utility analysis that has intensely occupied health economists. Almost thirty years of theoretical debate and empirical research has yielded no agreement. The conclusion that this agreement is absent actually is unanimous (Dol00, Dru97, Fro89b, Fro89d, Gol96a, Joh96, Len00, Nor92b).

Many different measuring procedures are available. They fall into two separate categories. First, there are the choice-based procedures: the standard gamble* (SG), the time trade-off (TTO) and the person trade-off (PTO). Second, there are the psychometric procedures, such as the paired comparison, the rating scale, the visual analogue scale and the magnitude estimation. The various procedures are in turn families of similar techniques that on closer inspection appear to differ in aspects such as the choice of the duration of the health states to be assessed in the SG, the duration of the remaining life expectancy in the TTO, the definition of the interval scale anchor points, the manner in which the health states are described and the precise wording of the questions. These are all aspects that can each influence the measurement results in their own way (Len00, Nor92b, Smi93).

The Washington Panel (Gol96a), the authors of the Canadian guidelines (CCO97), and Froberg and Kane (writers of four somewhat older overview articles to which referrals are still made (Fro89a-d)) view both the psychometric and the choice-based procedures as suitable for use in CUAs.

Others express a clear preference on theoretical grounds for the procedures based on choices, because these, in contrast to the psychometric ones, require people to make trade-offs between different arguments in their utility function, whereby they accord with economic thinking (Bra99, Bus00, Dol00, Joh96). The SG and the TTO are the most serious nominees in their eyes. But which of the two should be chosen? The debate on this among health economists is still going on (Dol00).

Both procedures have their problems. For both the SG and the TTO, the hypothetical choices put before respondents are not realistic (Joh96; compare 3.3.1). In addition, the measurement results from both procedures are influenced, albeit not in the same way, by a reluctance to give up life that varies from person to person and has, in reality, nothing to do with the relative assessment of health states (Bra99, Fow95). A further argument against the SG is that there are strong indications that, in practice, people often systematically violate the axioms of the expected utility theory on which the SG is based (Ble02). A more important drawback to the TTO is that the measurement results are sub-

---

* In the SG, respondents are given a choice between two alternatives: a particular chronic state, or a lottery with 'perfect health' and 'death' as possible outcomes. The question is at which chance for 'perfect health' the choice between the two alternatives is perfectly even. In the TTO, respondents are also asked to make a choice: to continue the rest of their life in a particular chronic state, or to live for a shorter period in perfect health. Here the question is which part of their life expectancy they are willing to give up in order to achieve the state of perfect health. The PTO is applied to the SAVE (see 3.3.1). Another possible formulation is: how many people in a particular chronic state would need to have their life extended by 1 year to make this option equal to extending the life of 100 people by this same 1 year? In the paired comparison, respondents must decide which states from a series of pairs of health states they prefer. In the rating scale, the visual analogue scale and the magnitude estimation procedures, they must indicate what value they attach to health states by assigning a number, marking a line, or by saying 'how many times better' (or worse) one state is compared to another (Dru97, Gol96).

ject to the respondents' time preferences (Bra99, Dol00, Fro89b, Joh96). Bleichrodt identified four forms of bias that play a part in both the SG and the TTO. These biases generally result in an overestimation with the SG, while they seem to cancel each other out in the TTO. The latter effect, however, is not guaranteed (Ble02). Virtually all authors agree that the third choice-based procedure, the PTO, is still insufficiently substantiated and developed. Additionally, a disadvantage for the PTO as a method for measuring the value of individuals' health states is that incorporating preferences about the distribution of health gain is inevitable.* However, the PTO does seem to be enjoying a rising degree of interest as a method for determining the social value of health care services output (Gre01, Men99, Nor99, Ube00b, Ube00c).

As far as research has been able to establish, the measurement results of the psychometric procedures seem to correlate only moderately with those of SG and TTO (Bra99). The values measured with the SG and the TTO mutually display a higher correlation, but the SG values tend to be somewhat higher than the TTO values (Ble97d, Bra99), which is in agreement with the expectations on theoretical grounds (Dol00). No one ventures general statements in the literature on the degree to which the measured values of the different procedures differ from each other. The Washington Panel concludes in 1996 that much more research is still needed into how the measurement results of the various procedures relate to each other under various circumstances (Gol96a).

This conclusion still holds six years later. The TTO appears to be gaining popularity among health economists, after the SG was long seen as the (certainly theoretically) superior method (Ble02). But there is no question of a general consensus, and uncertainty remains about the validity of health state valuations, even if they derive from a TTO procedure (Gre01, Len00). Bleichrodt, a self-declared TTO advocate, states in a recent article:

> "the existence of biases in SG and TTO utilities, the direction of which cannot always be predicted, highlights that the major challenge for health utility measurement is to develop utility measures and/or utility elicitation procedures that avoid or minimize the impact of biases on health utilities." (Ble02)

The uncertainty about the right procedure for measuring health state preferences also affects the health state classification systems recommended by health economists (see 3.2). The values attributed to health states in the various classification systems were obtained with different measurement procedures, both psychometric and choice-based, and for this reason alone they are not necessarily comparable (Bra99). In addition, they are dependent on the tenability of the multi-attribute utility models that lie at their foun-

* A version of the PTO specifically developed for the Global Burden of Disease Study was used in determining the disability weights for calculating DALYs (Mur97, Sto00a).

dation: the mathematical models that describe how the utility of the health states defined by the classification system relate to the utility of the various aspects or dimensions of the states (Dru97). Thanks to the use of these models, it is unnecessary to assess the utility of a large number of multidimensional health states. Instead it is enough to value the various levels of separate health state dimensions while the remaining dimensions are kept constant (Fro89a). This simplification of the valuation task has its price in the form of more or less stringent utility independence conditions. The least restrictive condition, belonging to the most complex model, requires that the utility of the various levels of each dimension (for example, the degree of mobility) is independent of the level at which the remaining dimensions (for example, pain or emotional conditions) are fixed (Fro89a). There is still discussion on whether this so-called decomposed approach to the valuation problem on which health state classification systems are generally based is preferable to the composite or holistic approach that requires the valuation of a large number of complete health states (Dol00). What is clear is that assessing a complete universe of possible health states without the simplifying assumptions of the multi-attribute utility models is a complex and time-consuming affair (Bra99, Dru97). Comparative empirical research into the validity of the decomposed and the composite approaches has not yet been performed (Dol00).

Chapter 4

Costs

If a CUA is aimed at supporting a societal decision about, for example, whether or not to fund a health care service from collective resources, then the CUA must in principle take into account all the costs that are associated with the service at issue. This means not just the costs that are involved in applying it (the direct health care costs: costs of buildings and equipment, material costs, staff costs), but also the costs that patients must incur in order to undergo the treatment. These include travel expenses, costs of child care and of special diets, costs of the time the patients and people in their environment devote to care and treatment (the direct non-health care costs), and the costs of the reduced 'productivity' in work and leisure time as a result of illness or death (the indirect or productivity costs).

So far there has been a large degree of consensus in CUA literature, albeit that the precise terminology and categorization of the types of costs remain the subject of discussion (Oos00). It is less clear how these kinds of costs must be estimated. For example, estimating the direct health care costs is far from simple. It will sometimes be possible to rely on existing prices and tariffs. However, these will often provide a poor reflection of the actual costs, because health care is a non-competitive and highly regulated social sector. In that case, it is necessary to estimate the costs. The judgement as to whether such a costs estimate is indeed necessary and, if so, just what degree of thoroughness and precision needs to be used, depends on many factors. These include the degree to which the outcomes of CUA appear to be influenced by these costs in a sensitivity analysis. It is not possible to formulate unambiguous decision rules for this (Luc96).

Long-standing controversies exist in particular with regard to productivity costs and costs incurred during life-years gained by medical treatment (Gar96, Hur00, Sto00b).

4.1 Productivity costs

Concerning the productivity of patients, there is fairly general agreement that the productivity gain resulting from a postponement of death needs to be taken into account in the denominator of the C/U ratio, that is in the health gain measured in QALYs. Productivity is part of the value that extending life represents to us. Double counting would occur when the productivity gain from extending a lifespan would also be factored into the numerator of the costs/QALY ratio, as negative costs (Gar96, Luc96, Wei97a). Weinstein pointed out that not all the costs of the death of a productive member of society are brought into the picture when calculating QALYs. Given that the effects on family and friends are generally ignored in CUAs ("perhaps an important limitation of the current practice"), he distinguished two kinds of costs that should be factored into the numerator, the cost component of the costs/QALY ratio: the friction costs, *i.e.* the costs incurred by replacing the deceased person in his or her workplace; and that portion of the deceased's income from which people other than the deceased benefited, such as "external consumption by others in society, financed by taxes and other transfers" (Wei97a). Weinstein did not provide a method for measuring these latter kinds of costs. Current guidelines for carrying out CUAs also ignore this (CCO97, CVZ99, Luc96).

The discussion within the literature concentrates on productivity loss in relation to morbidity. The principle that double counting should be avoided is also applicable here. However, it is less clear what it involves in this case. Different viewpoints are represented in the literature. The Washington Panel opted for an approach that agrees with what has just been outlined with respect to mortality. The Panel recommended that the entire influence that the illness exercises on the patient's quality of life, including the degree of productivity, should be incorporated in the denominator of the costs/QALY ratio (Luc96, Wei96). The Panel also distinguished, as did Weinstein regarding death, the friction costs and the consumption externalities that are carried by the employer, colleagues or the rest of society. Both these kinds of costs should be included in the numerator of the C/U ratio (Wei96, Wei97a).

Others maintained that this solution is not obvious and that it is also contrary to current research practice in that it assumes that people valuing health states take possible income losses into account (Bro97a, Mel99a). Should they in fact do this, then the valuation of health states would be dependent on the national level of social security (Bro97a). Rotterdam researchers contended that the friction costs, broadly interpreted as the sum of the productivity losses that occur as long as the ill employee is not replaced and the costs involved in coming up with adequate replacement (Koo94, Koo95), repre-

sent the actual productivity loss from a societal perspective (Bro97a, Bro97b). In this approach, the productivity costs resulting from illness are time-dependent and place-dependent. After all, the so-called friction period, the period needed to replace an ill employee, depends on the scarcity of labour. Thus the average friction period increased from 96 to 123 days in the Netherlands between 1995 and 1998 (Oos00). Other critics of the Washington Panel had little empathy with the Rotterdam friction-cost method and believed that it is preferable to include the ill employee's entire productivity loss in the C/U ratio numerator (Mel99a).

Reacting to criticism of their proposal to incorporate production losses resulting from illness in the QALYs, a few members of the Washington Panel conceded that their consensus on this point was fragile, weakly grounded in theory and provisional. But they maintained the Panel stand, "to avoid monetizing items that could be viewed as consequences of health-related quality of life" (Wei97a, Wei99). At the same time, they stated that more research is needed to produce better, workable and credible solutions (Gar99). However, one prominent Panel member believed that on second thoughts it would have been better to keep the productivity losses outside the QALYs, taking the view that "QALYs should and do reflect health apart from its uses" (Rus99).

Both Panel members and critics are convinced of the importance of the issue. Rotterdam researchers have demonstrated that the influence on the cost estimate of including or not including productivity losses can vary from zero to 50 per cent, depending on the type of services, the disease and the patient groups in question (Koo94).

4.2 Costs in life-years gained

Just as stubborn as the controversy about taking the productivity costs into account is that surrounding the costs incurred in the years of life gained through medical treatment (Wei97b). The Washington Panel distinguished between the medical costs associated with diseases that are either related or not to the intervention in question, and the non-medical costs. The Panel recommended that the related costs should be involved in the analysis. The Washington Panel reached no conclusion on whether or not to take into account the unrelated medical costs (Wei96, Wei97b). It is recommended that a sensitivity analysis should be carried out if necessary, to establish to what degree the unrelated costs influence the C/U ratio (Luc96). Non-medical costs should be kept out of the analysis (Wei96).

Others argued that, from a societal perspective, all the costs incurred in added years of life (related and unrelated medical and the non-medical) must be involved in the analysis (Mel99a, Mel99b). With their plea for calculating the non-medical costs (consumption minus productivity), they set themselves against the view that productivity gain from postponing death should be incorporated in the health gain measured in QALYs.

They pointed to the consequences of either not including, or not fully including, the so-called mortality costs. In groups that on balance consume resources, such as the elderly, this leads to a bias favouring services that primarily extend the length of life over services that mainly improve the quality of life (Mel97, Joh97). The opposite applies to groups that on balance produce more than they consume, such as the younger age groups (Mel00). However, the members of the Washington Panel held on to their viewpoint (Gar99, Wei99), and the controversy remained unresolved.

Apart from the complexity of estimating health care costs and the points of contention surrounding productivity costs and costs in life-years gained, there are also other problems that are less extensively discussed in the literature. These include those surrounding the costs of informal care, valuing the loss of unpaid work, and the costs of the lost leisure time of patients and those who support them (Bro00a, Koo98). Together with the complexity of estimating the direct health care costs, this makes estimating the cost-utility ratio numerator a delicate business that does not lend itself to straightforward resolution by unambiguous guidelines. Much must be left to the expertise and objectivity of the researchers in assessing exploratory sensitivity analyses, among others (Luc96).

The practice of CU research to date does not engender optimism on this point. A review of more than two hundred CUAs published between 1975 and 1997 showed that only 17% of the analyses accounted for any form of direct non-health care costs or time costs. The costs of patients' time were included in less than 10% and the productivity losses in 8% of the analyses (Sto00b). The authors observed that there was an appreciable variation in CUA cost research, in the choice of the included kinds of costs as well as in the manner in which they are estimated. This, the authors concluded, can have an important influence on the estimated C/U ratios. In addition, they determined that the source of valuation for cost estimates was often unclear or not reported. They concluded that "if CUAs are to play a major role in informing policymakers, more uniformity and transparency in cost estimation is needed as well as continued vigilance on the part of analysts, reviewers and journal editors" (Sto00b).

4.3 Marginal versus average costs

Apart from the methodological choices and problems mentioned earlier, there is yet another difficulty that generally limits the meaning of cost estimates in CUAs. In the foregoing (and in most CUAs) it is ignored that the health care costs associated with applying a particular service often depend on the scale on which it is applied, on the experience that is gained with it and on the degree to which the service is embedded in daily routines and logistics. In short, the costs are dependent upon various time-dependent and place-dependent factors (Ash00, Ger93). This insight is reflected in the health-

economic concept of the marginal costs, *i.e.* the added costs of producing one additional unit of the health output at issue (Gol96b, p 401).

The marginal cost concept derives its importance from the economic insight that in many situations it is not a dichotomous choice that is under discussion (that is, should a certain service be applied or not?). In general, the question is rather whether investment in a service, given a particular level of application, should either be increased or decreased. There are different thoughts about the significance that this concept must have for the design of CUAs. On the one hand, the Washington Panel made a recommendation that "costs in CEA should reflect the marginal or incremental resources consumed, rather than average costs, from a long-run perspective" (Luc96, p 209). On the other hand, pointing to "economies of scale, scope, indivisibilities, or learning-by-doing", the Panel stated that "in general, unless these effects are likely to be large, analysts can assume that the marginal costs of interest are constant". The Panel concluded with the rather empty advice that researchers "should consult the literature for appropriate adjustments" if the scale and learning effects are too great to ignore (Luc96, p.199). Rotterdam researchers attributed a much more limited meaning to the concept. They postulated that marginal costs "are pre-eminently relevant for specific research issues", but that "integral costs are to be preferred [...] when a generalization to national costs is necessary" (Koo98, Rut00b). They argued that the distinction between marginal and integral costs loses its meaning in the long term that is at issue with decisions at a national level (Oos00, Rut00b).

The Rotterdam standpoint is at odds with the opinion of Mooney, who maintained that the marginal analysis forms the core of the contribution that economic evaluation can make to the setting of priorities in health care. He suggested that there is, in general, a significant temptation to use average cost data, as these are often available, whereas data on marginal costs (at different levels of service application) will usually be lacking. But as Mooney put it, "the chances are very strong that the use of these average cost figures will lead to the costings being wrong". According to him, to assume that the relevant marginal costs will be different from the average costs is a much better starting point than to assume that they will be the same (Moo94b, p 31). In a similar spirit, Ashton and his colleagues argued that the relevant cost concept when priority setting must lead to budgetary shifts at the margins of health care programmes is that of the marginal costs. Unfortunately, the marginal costs are also difficult to measure because they are both context-dependent and volume-dependent (Ash00).

It is apparent from the views regarding marginal and average costs that there is no consensus as to precisely what should be understood as the health care costs in C/U ratios. Are these the average integral costs of a service in the long-term? But how long must the term be to allow the assumption that the costs that are fixed in the short run start to be variable and therefore relevant (Rut00b)? And how does this long term relate

to the possibility that the costs change through learning effects, scale effects, and organizational and technological developments?* Or should one focus on the average costs of a health care programme of a specific scale, with a specific form of organization, in a specific stage of technological development, at a specific point on the learning curve of the professionals involved? Or on the marginal costs of a particular extension of this programme?** And how are such time-dependent and place-dependent cost estimates to be extrapolated to other times and places (Oos00)? In any case, it is clear that users of cost-utility estimates cannot simply assume that the health care costs estimated in a CUA are also truly relevant in their own situation.

---

* Compare this with the remarks of Van Hout: "I believe much more work must be done to link the cost concept from economic evaluation studies to the cost concept in practice. […] In the future, economic analyses will need to take account of the programmes' scale and the time horizons." (Hou01)

** It is clear that marginal costs will often be at issue in decisions at the local level. But marginal costs will also often be of gcreater importance than average ones at the national level. After all, it is true for most health care services that their efficiency, except in the case of total ineffectiveness, is strongly dependent on the indications for which they are applied (CVZ01, Mul97, Nee00). An effective service is only rarely efficient or inefficient *per se*, but is mostly efficient in a particular (sometimes extremely limited) indication area, and less efficient outside it. Thus, a policy issue that will frequently arise in practice is whether it is sound from an efficiency point of view to fund a policy for a wider indication area when it is already funded for a limited indication, and thus to have it accepted on a larger scale. This is a policy question in which the relevant cost concept is that of the marginal costs.

Chapter 5

Discounting

People generally have what economists call a positive time preference. We would rather have positively valued resources now than later so that we may enjoy them in the meantime. In contrast, we would rather avoid unwelcome burdens as long as possible. In financial affairs, this is reflected in the interest rate.

Cost-utility researchers agree that this must have consequences for the valuation of future costs in a CUA. Costs must be expressed in one unit in order to add them up. Current monetary units are chosen for this. This means that future euros, dollars etc. must be converted to the present value by according them a lower value than current monetary units. This is called discounting.

But what about *QALYs* gained in the future? Are these also of less value than currently enjoyed QALYs? In the chapter on the theoretical foundations of CUA, the Washington Panel report indicated that discounting is "controversial" when health effects are concerned (Gar96). The authors accorded an important role to empirical research into the time preference that people generally show in their questionnaire answers and in their actual behaviour. Based on this research, they suggested that a certain degree of discounting is required, of costs and of health effects, but that there is substantial disagreement on whether the discount rate should be the same for market goods and for non-market goods. If there are no directly available market goods that could serve as a substitute for a non-market good such as health, then the discount rates do not have to be identical for all goods and services. According to the authors, powerful arguments could be put forward to bolster this proposition (Gar96).

Subsequently, in another chapter, the Panel offered a broader overview of the relevant literature and formulated its recommendations with respect to discounting (Lip96). Here they raised the often-heard suggestion that discounting effects and costs at the same rate would lead to a systematic and unjust undervaluation of the efficiency of preventive interventions. The authors swept this suggestion resolutely away, with a reference to a statement of the economists Fuchs and Zeckhauser: "self-respecting economists should not [...] use different rates because it is health that is being valued." According to Fuchs and Zeckhauser in 1987 (and the Washington Panel in their footsteps in 1996) if there is a reason to emphasize the value of preventive programmes, one should "adjust […] valuations of future benefits upward […] not [the] discount rate downward."

The rest of this chapter showed that there was in fact still heated discussion among economists on precisely whether there could yet be reasons to discount costs and health effects differently. In addition, the chapter offered an overview of the empirical research into individual time preferences, where interesting findings come to the fore that are relevant for, among other things, the assessment of preventive interventions:

- individual discount rates often lie outside the conventional range of 0 to 10%; 0% occurs very frequently, but rates *lower* than 0 and (much) higher than 10% also appear regularly
- discount rates tend to be lower when large-magnitude outcomes are being traded over time
- discount rates tend to be lower as the time interval to which the discounting applies gets longer
- discount rates for losses are typically lower than for gains.

As the authors of the chapter concluded, there is an abundance of behavioural evidence that indicates that the preferences of individual people are not in accord with the discounting of all possible goods at one particular rate.

Nonetheless, the Panel concluded its explanation with the recommendation that costs and health effects should be discounted at the same rate in the 'reference case analysis' (*i.e.* the methodologically standard CUA which the Panel propagated to promote the comparability of CUAs). But they added to this by stating that, since theory and empirical research offer no clarity about the relationship between individual time preferences and accepted discount rates, "sensitivity analyses based on models that allow health and costs consequences to be discounted at different rates may be conducted." The Canadian and Dutch guidelines for pharmaco-economic research contain similar recommendations (CCO97, CVZ99).

As expected, these recommendations do not signal the end of the discussion in the literature. Van Hout argued that there is quite a lot to be said for discounting costs and health effects at different rates (Hou98). He rejected the most important theoretical arguments that the Washington Panel had advanced in favour of discounting at the same rate, the consistency argument of Weinstein and Stason (Wei77) and the so-called Keeler-Cretin paradox (Kee83). From a societal perspective, contended Van Hout, costs need to be discounted on the basis of the expected increase of national income. On the contrary, the discount rate for health benefits should reflect expectations about increased efficiency in obtaining health benefits and the value that society attaches to these benefits. He concluded that it is well justified that the health benefits of, for example, a vaccination programme for children, be discounted at a lower rate than the costs, or not discounted at all, supposing that the decision makers' responsibility does not only cover the current generation but also future generations and that it is not plausible that all kinds of diseases will be treated more efficiently in the future than they can be treated today.

Other Dutch researchers also criticized the prevailing recommendations on discounting health gain (Bar99, Ble00, Bon01b, Bro00b). Bleichrodt and Brouwer wrote: "The decision to discount effects at the same rate as costs is arbitrary. The arguments in the literature for using one discount rate for effects and costs are not relevant to most policy decisions" (Ble00). The Dutch do not stand alone (Ana97, Goo96, Kra93). Anand and Hanson criticized the 3% discount rate that forms part of the DALY formula. They suggested (with a reference to several classical economic arguments against the idea that future utility or welfare should be discounted) that the only defensible argument in favour of discounting health gain is "the possibility that the world may end". Note, they went on to say, that "a 3% discount rate implies a 50% chance that the world will end in 23.4 years" (Ana97).

The discussion in the scientific literature on how to discount health gain correctly is thus still in full swing. This observation is an important one when it comes to the comparison of the efficiency of services with differing time relationships between investments and health effects (for example, preventive and curative services). The choice of the discount rate can make or break the efficiency of services, particularly of preventive ones (Bar99, Bon01b).

Chapter 6

Distributional effects

Uncertainties as to the correct way to quantify the cost-utility ratio could be consequential for the distribution among the population of the health gain that could be generated with a given collective health care budget. A health state assessment that is insensitive to the disease-specific effects of a particular treatment, a random sample from the general population that is insufficiently able to value a subtle difference between health states before and after a treatment, a valuation method that under-values the difference in utility of two states because of respondents' reluctance to give up life: all these, and many other, factors can result in CUAs that either underestimate or overestimate the health gain brought about by a particular health care service. Thereby, specific categories of patients may be disadvantaged, to the extent that the results of CUAs indirectly (through the meaning ascribed to CUAs by the responsible decision makers) lead to decisions to refrain from collectively financing services from which these patients could benefit.

How the uncertainties and inadequacies in the CUA methodology affect the distribution of health gain cannot be said in general for many of them. The distributional effects are dependent on specific combinations of the characteristics of populations, diseases, treatment indications and services, on the one hand, and the methodological choices made in CUAs, on the other hand. But there are also CUA characteristics that have attracted attention in the literature precisely because it is (or seems to be) relatively easy to figure out which distribution effects emanate from them.

There has thus been exhaustive discussion of the fact that, all other things being equal, the elderly have (because of their shorter life expectancy) fewer QALYs to gain than the young. Distribution decisions based on CUAs thus favour the young over the

elderly. There is diversity of opinion on whether this is justified (Cal87, Cal90, Dan88, Eme89, GR86, Har87, Keu91, McK98, Men90, Mus87, Raw89, Rig86, Sas01, Smi87, Tsu00, Wag00, Wil92). A similar effect occurs in the quality dimension among people with chronic diseases and permanent functional limitations. With a specific number of years added to their life, these people have fewer QALYs to gain than those without such diseases and limitations (Ana97, Men99, Nor99, Nor01). It has been pointed out that CUA does not take account of what is called the rule of rescue (Jon86): the belief that neither cost nor effort must be spared when someone can be rescued here and now from impending death (Had91, Men99). Furthermore, it is argued that because CUAs measure health gain in QALYs, they unjustly ignore the fact that representatives of the general population, all other things again being equal, generally rate a particular change in the health state of a seriously ill person more highly than the same change in the condition of a largely healthy person (Dan93, Men99, Nor99, Sas01).

Less frequently discussed, but no less true, is the observation that, apart from age, other demographic characteristics such as gender, race and socio-economic status are also powerful predictors of the results of a medical treatment, so that a distribution of resources based on CUAs is at odds with the pursuit of equal access to health care for all, without regard to these characteristics (Bon91, Bon93). Distributional effects also stem from how the aggregation problem is solved in CUAs. By adding up the health benefits, measured in QALYs, of different people, the CUA embodies a specific, and debatable, answer to questions such as how to weigh the saved lives of a few against the cured sore throats of many (Dan93, Sas01).

Apart from these implications of how health gain is measured in CUAs, there are also distributional effects that originate on the cost side of the cost-utility ratio. Several types of costs are dealt with in different ways, which may cause significant distributional effects. Examples include the discounting of the lost leisure time of patients and informal care providers, the valuation of unpaid work, and the manner of dealing with age-dependent and gender-dependent differences in earnings (Rus96).

It is clear that CUA, applied as an aid to reach a maximization of QALYs gained with a given health care budget, brings with it all kinds of more or less visible implications for the distribution of the health gain achieved. A distribution of the collective health care budget on the basis of cost-utility ratios, in which all QALYs are weighted equally, irrespective of the characteristics of those who benefit, leads to a specific distribution of the health gain achieved, that not necessarily has to be a just distribution.

In the health economic literature, the view that ‘QALY maximization’ may be at odds with ‘justice’ is widely accepted. It is recognized that equity-efficiency trade-offs are regularly involved in the allocation of resources in health care: efficiency must, to a certain degree, be sacrificed to justice and *vice versa*. The question as to how these trade-

offs can best be approached has in recent years received a steadily growing amount of attention in the literature. However, for the time being a satisfactory answer is not yet in sight (Ble97a, Pol02, Sas01, Wag91, Wil00). On the one hand, efforts have been made to come up with weights to weight QALYs differently, depending on to which category of people they accrue (young or old, seriously ill or largely healthy people, etc.). However, it is acknowledged that the research needed to be able to base these weights on the preferences of the general population is still in an embryonic stage (Wil00). On the other hand, it is suggested that, certainly in economic evaluation research in the short run, it should be consistently shown just which subpopulations (defined, for example, in terms of age, gender, ethnicity and socio-economic status) would benefit from a health care service. This will allow decision makers to assess the distributional effects of alternative spending options (Sas01).

An entirely different approach is implied in the previously mentioned saved young life equivalent, or SAVE (Nor92a). The SAVE quantifies the social value that is accorded to a particular health improvement in a person with a specific health condition, a specific age and other possible characteristics. The question to be answered is: how many such improvements in this category of persons counterbalance the total health recovery of one young person who would otherwise have suffered a certain and quick death (cf. 3.3.1)? Thus, the SAVE incorporates both efficiency and equity considerations, leaving it open to criticism from those health economists who precisely see the clear distinction between these two types of considerations as one of hallmarks of a good health economic analysis (Wil00). However, those who actually agree with Menzel that any attempt to quantify health gain in QALYs "will remain persistently suspicious" if the allocation decisions that constitute the context of use of QALYs are not taken into account (Men90), will not be impressed by this criticism. From their point of view, it is – after all – entirely credible that it will turn out to be simply impossible to quantify efficiency in health care plausibly, independently of equity considerations.

Chapter 7

Conclusion

"The QALY concept has already been applied for more than 25 years and is described in countless scientific publications. Scientists did initially express their reservations about the quality of QALYs, but in the interim this concept has been thoroughly substantiated in such a way that QALYs are used worldwide in policy models." (Pol02)

"QALYs seem to be the most widely used method for capturing both quality and quantity of life. However, the QALY concept is controversial […]." (Gre02)

Is efficiency a usable criterion for determining the composition of the basic health care benefit package? That was one of the questions the then Dutch Minister of Health, Welfare & Sport put before the Health Council in her request for advice. The Minister referred to the Public Health Care report of the Netherlands Scientific Council for Government Policy. In that report, the Council gave a specific interpretation to the efficiency criterion and drew a rosy picture of its usefulness. The Council believed that the cost-utility of services, expressed in costs per Disability Adjusted Life Year (DALY) to be gained, should be the criterion in preventive and curative health care for the allocation of collective resources. The Council remarked that "awkward choices do remain in the proposed methodology", but noted that "In some respects, the matter is one of resolving technical issues"– issues that should not be a reason for not using the cost-utility criterion.

In the preceding chapters of the present report, these "technical issues" have received full attention. The result will not come as a surprise for health economists, who

are generally well familiar with the problems observed in their scientific literature. However, it does provide food for thought for the decision makers who, possibly contrary to knowing better, sometimes seem to hope that in cost-utility analysis they have at hand a relatively simple instrument that will relieve them of many difficult considerations. The CUA and the Quality Adjusted Life Year (the QALY, the measure of health gain distinguishing this form of cost-effectiveness analysis from others) still appear to be open to discussion in several respects more than 25 years after their introduction.

Standardization is not a solution to this problem. The often-heard call for standardization is prompted by precisely the conclusion that appears inescapable in the light of the previous chapters, namely, that there are methodological differences between CUAs that have meaningful effects on the outcomes and that make the comparability of C/U ratios debatable (CCO97, CVZ99, Gol96b). But standardization does not necessarily lead to the right methodology. There is no reason to assume that one particular set of methodological choices would lead to the same ranking of services by cost-utility as another, equally defensible, set. Moreover, even if there was agreement, this would still not prove that the ranking is correct. Thus, new rounds of discussion have always followed initiatives towards standardization.

What does this mean for the usefulness of the C/U ratio and, returning to the Minister's actual question, for the applicability of the efficiency criterion in defining a basic health care benefit package to be financed collectively? Having arrived at this point, it is useful to distinguish anew between the various aspects of the efficiency concept that have been raised in this report. The problems we encountered in the previous chapters are different in nature and require different approaches. They are reviewed here afresh, but this time in reverse order, so that we conclude with the issue of the QALY.

Discounting

Is the value of health gain dependent on the moment, now or in the future, at which it is achieved? If not, then how must future gain be weighed against current gain? These questions are discussed extensively in the health economic literature. A unanimous answer has not emerged.

The question of whether the efficiency of a health care service is dependent on the manner of discounting future health gain can be determined relatively easily in a sensitivity analysis. In practice, the dependence is particularly large in preventive services and public health programmes (Kra93). As long as services with comparable time relations between costs and health gain are compared, the uncertainty about discounting will produce few problems. However, efficiency comparisons beyond these limitations (for example, a direct quantitative comparison of the efficiency of preventive and curative

services) require clarity regarding the correct manner of discounting. Without scientific consensus, only the government can create this clarity (Kra93, Bro00b).

Should the uncertainty surrounding the discounting of health gain continue, it would mainly be an impediment to a comparison of C/U ratios across the entire breadth of health care. The comparison of C/U ratios and cost-effectiveness ratios within sectors of health care, where there is a certain homogeneity in terms of the time relationship between costs and health gain, will generally be less troubled by the discounting problem.

Costs

The problems involved in quantifying costs have by no means been covered exhaustively in this report. It is clear, though, that differing ways of approaching the various types of costs (for example direct health care costs, productivity costs, costs in life-years gained and informal care costs) can lead to important differences in estimating costs. To a certain extent, the use of sensitivity analysis makes it possible to ascertain to what degree an efficiency estimate is dependent on variations in costs. But in this case, the sensitivity analysis is much less simple than when it only concerns the effect of different ways of discounting future health gain. After all, there may be a lot of different cost variables at issue, including variables that are interdependent (for example, costs of informal care and direct health care costs) and variables that vary over a range that is not obvious *a priori.* Moreover, the methodology for estimating the cumulative influence of the uncertainty in various, partly interdependent, variables is still under development (Bri95, Bri97, Bri98a, Bri98b, Bri00, Gar00, Hun98b, Man96).

Problems in estimating costs trouble every type of efficiency analysis, but they turn out more awkward the higher the ambition of the analysis is set, thereby enlarging the scale of relevant cost types. From the general societal perspective, in principle all costs are relevant. However, there is no agreement among health economists on all points as to which choices and definitions of costs follow from that principle. Therefore, it seems inevitable that the government will itself cut some of the knots, if it wants to involve the efficiency criterion in its decision making on defining the basic package (Boe02). Will it consider as relevant the productivity losses that result from illness? If so, what does it understand by this: are the integral productivity losses or the (labour market dependent) friction costs at issue? In estimating productivity losses, is the use of age-specific and gender-specific average wages appropriate? And as far as the costs in added life-years are concerned: is the government interested in these costs and if so, is it interested only in the related medical costs or also in the unrelated ones, or even in the non-medical costs? There is much to be said for and against all these options, and that is exactly what

has happened amply in the health economic literature. It seems that making choices is more the task of the government than of science.

Health gain

When we measure the output of health care in terms of 'health gain', we do not incorporate everything that makes health care valuable. There should be room to consider other aspects of the output when making decisions on the definition of the basic package. However, that does not distract from the central role of health gain, as a component of efficiency and also, incidentally, as an independent criterion.* How should health gain be defined? Changes in lifespan and changes in health state are both relevant, as everyone would agree. It is less self-evident that these changes could or should be valued in one generally applicable dimension. The most current universal measure to quantify health gain, the QALY, is fraught with numerous difficulties. There are problems in integrating the 'quality' and 'quantity' of life, in aggregating the QALYs of different people, in describing health states in a standard manner, in formulating a theory about health state preferences, in developing procedures for measuring the preferences, and in the choice of the valuators.

On the one hand, these problems give rise to questions about the interpretation of the QALY: is the QALY actually a useful measure, is it an entity to which a clear meaning can be attached? On the other hand, these problems can be approached partly from a quantitative angle. In principle, it is possible in a sensitivity analysis to determine to what degree the estimation of health gain in QALYs depends on uncertainties in estimating the quality adjustment factors. But that is not at all straightforward, since many different health states may play a role. Besides, a good assessment of the effect on the resulting cost-utility ratio requires a methodologically complicated, multivariate sensitivity analysis. In such an analysis, the uncertainties in the estimated duration of the periods spent in the various health states must also be involved, as must be, ultimately, the uncertainties in the assessment of the costs. Furthermore, it is difficult to see how the more fundamental theoretical problems, such as those in integrating life quality and life duration as well as in the aggregation of health benefits of different people, could be dealt with in a quantitative sensitivity analysis.

Nevertheless, it is conceivable that the problems associated with the quality adjustment of life duration are of little relevance in particular situations because efficiency is determined to a large degree by the postponement of death and very little by the quality

* Suppose that two treatments are available for an acute, fatal disease. Treatment A extends life by one year and costs one hundred euros. Treatment B yields a ten-year life extension, but costs one hundred thousand euros. The efficiency of treatment A, expressed in direct medical costs per gained life-year, is then a hundred times higher. However, treatment B is actually preferable from an effectiveness point of view.

of life. But one may easily go astray here. For example, in the case of treatment with a cholesterol synthesis inhibitor ('statin') in the prevention of coronary heart disease, it seems apparent that efficiency is strongly determined by the postponement of death and only to a lesser degree by the reduction of morbidity (Bon99). The economic analysis to support the development of the Dutch consensus on the indications for treatment with statins, also primarily occupied itself with the costs per life-year gained (CBO98). Nonetheless it appeared in a later CUA that the costs of statin treatment per gained life-year differed by 20% from those per gained QALY (Pro00). In addition, a minor difference between the costs per life-year and per QALY gained could also mean that significant effects on the quality of life have been missed in the CUA. For example, the Washington Panel (Gol96) pointed to the discrepancy between the large difference shown in the Australian CUA of breast cancer screening discussed in 3.2 (Hal92) and the 5.6% difference found in a Dutch CUA (Kon91).

Research demonstrating how differences in the methodology of life-year quality adjustment work out in cost-utility estimates, is not amply available. Apart from the previously mentioned examples (Hal92 in 3.2, Gab99 in 3.3.2), research by Hornberger and colleagues (Hor92) showed that hemodialysis cost-utility ratios can vary from $35,000 to $45,000, depending on the method of health state valuation. However, there are numerous research examples showing that health state valuations can differ markedly depending on the methodology used (Bel01, Fry93, Nea95, Neu00b, Nor92b, Rea84). It is, therefore, plausible that the problems associated with life-year quality adjustment can not be dismissed as being quantitatively inconsequential. After all, many services (for example, for chronic diseases such as gastrointestinal complaints, disorders of the locomotor apparatus, lung disorders and mental disorders) are aimed primarily at improving the health state and not at extending life. In addition, many life-prolonging services also have an important effect on the quality of life. Thus, for example, a Dutch study showed that the costs of lung transplantation per QALY gained were more than 20% lower than those per life-year gained (Enc97).

All in all, the conclusion that the QALY is not a generally applicable health gain measure that can be relied upon when comparing the efficiency of differing health care services, seems to be warranted.*

Alternatives to the QALY

Is there another universal, one-dimensional measure of health gain ahead that could take the place of the QALY in the CUA? To date, three candidates have been pushed to the fore in the literature: the DALY, the Healthy Year Equivalent (HYE) and the saved young life equivalent (SAVE).

The DALY was developed primarily as a measure enabling the international comparison of the burden of disease, but was also embraced as a measure of health gain in CUAs (Mur96, Mur97). It is not much more than a 'reversed QALY'. The DALY combines lost years as a result of premature death with weighting of lived years for the degree of disability. The result is a measure of health loss instead of a measure of health gain, as the QALY. The problems are identical.

The less well-known HYE, based on the valuation of lifetime health profiles instead of isolated health states, is intended to offer a solution to particular QALY problems (see 3.1). The theoretical superiority of the HYE to the QALY seems obvious, although it is not undisputed (Rie98). However, the measurement of HYEs is not straightforward from a practical viewpoint and it has to date been little used in empirical research. Nonetheless, the HYE approach still seems to warrant further exploration (Lle02).

The SAVE offers a more radical alternative to the QALY (see 3.3.1). This measure quantifies the social value of health states changes with explicit reference to the type of distribution decisions in support of which the health gain is measured. The SAVE escapes many of the problems identified in Chapter 3, and also seems to offer a way of handling, in a direct and insightful manner, the distributional aspects that are connected to the valuation of the health care services' output. Even more than the HYE, the SAVE seems to be a measure that is worth a closer look in the Dutch context.

---

* This report is not alone in its reserved conclusion. Indeed, precisely the opposite is the case. It is fairly commonplace in the methodologically oriented literature about CUAs not to speak highly of QALY credibility. For example, the Washington Panel wrote the following about the different ways of valuing health states: "This diversity in how preference weights are gathered markedly constrains the ability to compare analyses credibly where the effectiveness measure is presented in QALYs" (Gol96, p 119). The health economists Drummond and Torrance warned in an article against the misplaced suggestion of comparability that can be derived from so-called league tables where cost-utility ratios emanating from different studies are simply lumped together. According to them, "[…] the results of economic evaluation studies should be presented in a disaggregated form as well as in the form of cost-effectiveness ratios. […] the consequences of interventions in terms of their physical effects should be reported separately from utility weights used to value them." (Dru93)

There is still another promising approach to the problem of valuing health gain. However, this willingness to pay (WTP) approach was not covered in the foregoing because it falls outside the domain of CUA. The WTP, also sometimes named contingent valuation, values health gain in terms of money and can thus be used in cost-benefit analysis, which is rated highly by health economists because it is strongly grounded in the welfare economic theory (Dru97). In particular, the WTP question which asks respondents to indicate how much extra insurance premium they would be prepared to pay to add a particular service to the benefit package (the ex ante insurance based WTP), expresses with an agreeable, frank directness the unpleasant question that actually lies at the core of the allocation of (collective) resources to health care services: what is it worth to us? The WTP approach also has the advantage that dimensions of the health care output other than health gain are relatively easy to incorporate in the assessment (Bir99, Lee97, Neu94). Undoubtedly, the WTP also has significant problems (Ble01, Klo99, Ols01). Nonetheless, anyone who looks over the development of health economic evaluation research will wonder whether the aversion to valuing health in terms of money, which has played such an important role in the (over)valuation of CUA to the detriment of cost-benefit analysis (Gar96, p 28; Wei80, p 240), has not been a bad counsellor.* The WTP has for some years been enjoying increasing attention in the international health economic literature (Bir99, Dru97, Klo99, OBr96). Re-evaluation also looks imminent in the Netherlands.**

No matter how attractive the SAVE and the WTP, in particular, appear to be, none of the four listed approaches to valuing health gain can currently (or possibly even over a period of some years) be seen as a genuine alternative to the QALY. This means that a credible way of measuring health gain with one generally applicable measure is not available in the short term. Nonetheless, the government could choose to embrace one specific methodology for measuring health gain in QALYs. But, given the nature of the problems highlighted in the literature on QALYs and the enormous social pressure under which decisions on the distribution of collective health care resources must be made, this appears to be a strategy offering little chance of success. It will not be easy for the government to defend itself against the charge that the QALY is in essence “a technical solution to a political problem” (Car91).

---

* The same applies to the argument that the WTP could be influenced in an undesirable way by the respondents’ income (Ble01, Gar96, Ols01). As far as a study would show a relation with the income level, the WTP could indeed simply be standardized to any desired income level with the help of regression analysis.

** For example, contrary to the well-known book by Drummond and colleagues (Dru97), the WTP approach is barely (Uyl00) or only briefly (Blei01) covered in two recent Dutch books on health economic and health economic evaluation research. However, even these books does not exclude a more prominent future role for the WTP.

How to proceed with efficiency?

For the time being, there seems to be no choice but to accept that there are no generally applicable technical resources available for comparing the efficiency of health care services. The frequently multifarious effects of services on the health state, the effects on life duration and the costs will all need to be considered by the decision making bodies themselves when making their decisions. In addition, justice must be done to yet other considerations alongside efficiency.* These considerations include effectiveness, the effects on those other than the patient and on society as a whole, other aspects of the services' output, the degree to which (the efficiency of) a service could be further developed, the available budget, practical considerations, distributional justice in relation to dimensions such as the seriousness of the disease, prognosis, dependency and age, and other ethical and legal aspects. Moreover, the issue under consideration will often not be simply whether to fund or not fund services, but rather the specification of the indications by which the services should be funded.** These indications are continuously subject to change through new scientific and technological developments as well as the emergence of rival or additional interventions.*** The final piece of decision making on the distribution of scarce health care resources ultimately lies in the hands of care providers that need to decide whether the agreed indications for funding are present in individual patients.

How should one deal with this overwhelming complexity? Clearly, we are confronted with a major problem in terms of feasibility, but above all in terms of legitimacy. It is obvious from the international discussions and the experiences of the past 15 years that debates on distributional principles do not suffice. Alongside principles, it is essential to have the active participation of all involved parties and experts, a clear division of responsibilities and transparent decision making and dispute resolution procedures. All of these are needed in order to resolve differences of opinion about priorities in health

---

* This is generally not contested in the health economic literature (Ble01, Bro00a, Bus00, Rus96, Wei97b). Hurley pointed out that health economists often plead this in defence when they are confronted with criticism of their desire to incorporate all the important effects of health care into the QALY. They respond that "the results of economic evaluations are only one piece of information intended to aid decision making and that other relevant considerations will enter via other avenues at the time a decision is taken." However, Hurley argued, if this is true then much of the attraction of being able to rank all possible health care services in one cost-utility dimension evaporates. After all, the results of the efficiency analysis will need to be combined in the decision making process in some way with other types of data. And if this is so, he continued, then "why should the entire structure of the economic evaluation be distorted so as to obtain a single number at the end of the analysis?" (Hur98).

** Cf. the final footnote in Chapter 4.

*** The same dynamic also leads to rapid changes in what should be the alternative service with which a service to be judged on its efficiency should be contrasted (for example, in a CUA)(Ger93, Rut00a). Even just the rapid changeability of the contrast to be studied seems to be an important obstacle on the road to a prominent role for the laborious, methodologically demanding and data-intensive CUA in delineating a collectively financed health care benefit package.

care and to stimulate a collective social learning process (Bur98, Dan93, Dan97, Dan00, Day98, Ham97, Hol98, Kle98, Leg02, Nee00).

Demarcating well-defined areas of health care within which priority setting must come to a conclusion appears to be one of the necessary mechanisms for making the complexity manageable. Just how these areas must be defined is another matter. Examples may include care for the mentally disabled, ambulatory mental health care, prevention of cardiovascular disease, and solid tumor oncology. It seems clear, though, that a certain degree of homogeneity of types of problems and a manageable body of relevant knowledge and experiences are important preconditions for successful decision making.*

Nonetheless, even within these areas, achieving agreement on the indications by which services must be funded will be an enormous task. An aid such as the QALY will possibly provide good service here. After all, it is natural to suppose that the smaller the diversity of the compared services, health problems, patient populations and health gains, the less weight will be carried by the interpretation problems embodied in the QALY (Rus96). That is certainly the case when one is comparing the efficiency of applying the same service in various phases of the same disorder. One can think of the comparison between the efficiency of an operation either earlier or later in the development of prostate complaints, or of the use of cholesterol synthesis inhibitors depending on the risk of coronary heart disease. In efficiency comparisons in such homogeneous domains, calculating QALYs may not even be necessary because specific outcome measures, such as life-years or disease-free life-years gained, suffice. In addition, the SAVE and the WTP may also be helpful in making decisions within the defined areas of health care.

However, the legitimacy of the decision making in a particular area of health care will need to rest primarily with a carefully designed decision making procedure. In this procedure, parties that understand that health care area from the inside (especially care providers and patients) must participate actively. CUAs, and in the longer term possibly also cost-benefit analyses, can be aids to reaching an overview and agreement in this decision making process.

* Naturally, the other side of this reduction in complexity is that in this way the distribution of resources among the various areas of health care is not dealt with. This allocation appears to be an inalienable responsibility of the national government.

Literature

Aar96 Aaronson NK, Cull AM, Kaasa S, e.a. The EORTC modular approach to quality of life assessment in oncology. In: Spilker B, editor. Quality of life and pharmacoeconomics in clinical trials. 2e editie. Philadelphia: Lippincott-Raven, 1996: 179-89.

Ana97 Anand S, Hanson K. Disability-adjusted life years: a critical review. J Health Econ 1997; 16: 685-702.

Arn99 Arnesen T, Nord E. The value of DALY life: problems with ethics and validity of disability adjusted life years. Br Med J 1999; 319: 1423-5.

Ash89 Ashmore M, Mulkay M, Pinch T. Health and efficiency. A sociology of health economic. Buckingham: Open University Press, 1989

Ash00 Ashton T, Cumming J, Devlin N. Priority-setting in New Zealand: translating principles into practice. J Health Serv Res Policy 2000; 5: 170-5.

Bad01 Badia X, Roset M, Herdman M, e.a. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. Med Decis Making 2001; 21: 7-16.

Bar98 Barendregt JJ, Bonneux L. Degenerative disease in an aging population. Models and conjectures (Proefschrift). Rotterdam: Erasmus Universiteit, 1998.

Bar99 Barendregt JJ, Bonneux L. The trouble with health economic. Eur J Public Hlth 1999; 9: 309-12.

Bel01 Bell CM, Chapman RH, Stone PW, e.a. An off-the-shelf help list: a comprehensive catalog of preferences scores form published cost-utility analyses. Med Decis Making 2001; 21: 288-94.

Ber73 Berg RL, editor. Health status indexes. Chicago: Hospital Research and Educational Trust, 1973

Ber76 Bergner M, Bobbitt RA, Kressel S, e.a. The Sickness Impact Profile: conceptual formulation and methodology for the development of a health status measure. Int J Health Serv 1976; 6: 393-415.

Ber89 Bergner M. Quality of life, health status, and clinical research. Med Care 1989; 27, Suppl: S148-56.

Ber98 Berg M, Bezemer NM, van den Burg M. Normatieve aspecten van richtlijnen en kosteneffectiviteitsanalyses. Deelrapport 2 KNMG-project Passende medische zorg. Utrecht: KNMG, 1998.

Ble95 Bleichrodt H. QALYs and HYEs: Under what conditions are they equivalent? J Health Economics 1995; 14: 17-37.

Ble96 Bleichrodt H. Applications of utility theory in the economic evaluation of health care (Proefschrift, Erasmus Universiteit Rotterdam). Ridderkerk: Ridderprint, 1996.

Ble97a Bleichrodt H. Health utility indices and equity considerations. J Health Economics 1997; 16: 65-91.

Ble97b Bleichrodt H, Quiggin J. Characterizing QALYs under a general rank dependent utility model. J Risk Uncertainty 1997; 15: 151-65.

Ble97c Bleichrodt H, Wakker P, Johannesson M. Characterizing QALYs by risk neutrality. J Risk Uncertainty 1997; 15: 107-14.

Ble97d Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of qalys. J Health Economics 16, 1997: 155-175.

Ble00 Bleichrodt H, Brouwer W. Disconteren. In: Rutten-van Mölken MPMH, Van Busschbach JJ, Rutten FFH, editors. Van kosten tot effecten. Een handleiding voor evaluatiestudies in de gezondheidszorg. Maarssen: Elsevier gezondheidszorg, 2000: 123-9.

Blei01 Bleichrodt H, Koopmanschap M. Economische evaluatie. In: Lapré R, Rutten F, Schut E, editors. Algemene economie van de gezondheidszorg. Maarssen: Elsevier gezondheidszorg, 2001: 273-94.

Ble02 Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. Health Econ 2002; 11: 447-56.

Boe98 Boers M. Kunst van doelmatig onderzoek (Inaugurele rede). Amsterdam: Vrije Universiteit, 1998.

Boe02 Boer A. Onderzoek op maat. Een verkenning van factoren voor het gebruik van Technology Assessment. (Proefschrift). Rotterdam: Erasmus Universiteit, 2002.

Bon91 Bonsel GJ. Methods of medical technology assessment with an application to liver transplantation (Proefschrift). Rotterdam: Erasmus Universiteit, 1991.

Bon92 Bonsel GJ. QALY's: normatieve toepassing van gezondheidsmeting. In: Gunning-Scheepers LJ, Mootz M, editors. Gezondheidsmeting. Houten: Bohn, 1992: 122-33.

Bon93 Bonsel GJ. Kosteneffectiviteitsanalyse: over wikken en beschikken. In: Dillmann RJM, van Leeuwen E, Kimsma GK. Ethiek in de medische praktijk. Utrecht: Bunge, 1993: 167-78.

Bon99 Bonneux L, Barendregt JJ. Ook met de consensus over verlaging van de serumcholesterolconcentratie blijft primaire preventie duur. Ned Tijdschr Geneeskd 1999; 143: 772-5.

Bon01a Bonsel G. Wat heet beter? (Inaugurele rede). Amsterdam: Vossiuspers UvA, 2001.

Bon01b Bonneux L, Birnie E. The discount rate in the economic evaluation of prevention: a thought experiment. J Epidemiol Community Health 2001; 55: 123-25.

Bor95 Borst-Eilers E. Beleidsbrief Medische Technology Assessment (MTA) en doelmatigheid van zorg. Tweede Kamer, 1995-96, 24126, nr 9.

Bos00 Bosch JL, Hunink MGM. Comparison of the Health Utilities Index mark 3 (HUI3) and the EuroQol EQ-5D in patiënts treated for intermittent claudication. Qual Life Res 2000; 9: 591-601.

Boy90 Boyd NF, Sutherland HJ, Heasman KZ, e.a. Whose utilities for decision analysis? Med Decis Making 1990; 10: 58-67.

Bra99 Brazier J, Deverill M, Green C, e.a. A review of the use of health status measures in economic evaluation. Health Technol Assessment 1999; 3(9): 1-164.

Bri95 Briggs A, Sculpher M. Sensitivity analysis in economic evaluation: a review of published studies. Health Econ 1995; 4: 355-71.

Bri97 Briggs A, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. Health Econ 1997; 6: 327-40.

Bri98a Briggs A, Gray A. The distribution of health care costs and their statistical analysis for economic evaluation. J Health Serv Res Policy 1998; 3: 233-45.

Bri98b Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. Health Econ 1998; 7: 723-40.

Bri00 Briggs A. Handling uncertainty in cost-effectiveness models. Pharmacoeconomics 2000; 17: 479-500.

Bro93 Broome J. Qalys. J Public Economics 1993; 50: 149-67.

Bro95 Brock DW. Justice and the ADA: does prioritizing and rationing health care discriminate against the disabled? Social Phil & Policy 1995; 12: 159-85.

Bro97a Brouwer WBF, Koopmanschap MA, Rutten FFH. Productivity costs measurement through quality of life? A response to the recommendation of the Washington panel. Health Econ 1997; 6: 253-9.

Bro97b Brouwer WBF, Koopmanschap MA, Rutten FFH. Productivity costs in cost-effectiveness analysis: numerator or denominator: a further discussion. Health Econ 1997; 6: 511-14.

Bro00a Brouwer WBF, Koopmanschap MA. On the economic foundations of CEA. Ladies and gentlemen, take your positions! J Health Econ 2000; 19: 439-59.

Bro00b Brouwer W, van Hout B, Rutten F. A fair apporach to discounting future effects: taking a societal perspective. J Health Serv Res Policy 2000; 5: 114-8.

Bro01 Brouwer WBF, Rutten FH. Health Econ. A bridge over troubled water. Eur J Public Health 2001; 11: 234-6.

Bur98 van den Burg M, ter Meulen RHJ. Prioriteiten binnen de gezondheidszorg. Een vergelijkende studie naar normatieve overwegingen van prioriteitenstelling in Zweden, Engeland en Nederland. Deelrappport 1 KNMG-project Passende Medische Zorg. Utrecht: KNMG, 1998.

Bus00 van Busschbach JJ. Waarderend kwaliteit van leven-onderzoek. In: Rutten-van Mölken MPMH, van Busschbach JJ, Rutten FFH, editors. Van kosten tot effecten. Een handleiding voor evaluatiestudies in de gezondheidszorg. Maarssen: Elsevier gezondheidszorg, 2000: 77-94.

Cal87 Callahan D. Setting limits. Medical goals in an aging society. New York: Simon & Schuster, 1987

Cal90 Callahan D. What kind of life. The limits of medical progress. New York: Simon & Schuster, 1990

Car89 Carr-Hill RA. Assumptions of the QALY procedure. Soc Sci Med 1989; 29: 469-77.

Car91 Carr-Hill RA. Allocation resources to health care: is the QALY (Quality Adjusted Life Year) a technical solution to a political problem? Int J Health Serv 1991; 21: 351-63.

CBO98 Centraal Begeleidingsorgaan voor de Intercollegiale Toetsing (CBO). Behandeling en preventie van coronaire hartziekten door verlaging van de plasmacholesterolconcentratie. Consensus Cholesterol, tweede herziening. Utrecht: CBO, 1998.

CCO97 Canadian Coordination Office for Health Technology Assessment. Guidelines for economic evaluation of pharmaceuticals: Canada. 2e editie. Ottawa: CCOHTA, 1997.

Cul98 Culyer AJ. How ought health economists to treat value judgments in their analyses? In: Barer ML, Getzen TE, Stoddart GL, editors. Health, health care and health economic: perspectives on distribution. Chicester: John Wiley & Sons, Ltd, 1998: 363-71.

CVZ9 College voor zorgverzekeringen. Richtlijnen voor farmaco-economisch onderzoek. Amstelveen: CVZ, 1999.

CVZ01 College voor zorgverzekeringen. Het basispakket: inhoud en grenzen. Amstelveen: CVZ, 2001.

Dan88 Daniels N. Am I my parents' keeper? An essay on justice between the young an the old. New York: Oxford University Press, 1988.

Dan93 Daniels N. Rationing fairly: programmatic considerations. Bioethics 1993; 7(2/3): 224-33.

Dan97 Daniels N, Sabin J. Limits to health care: fair procedures, democratic deliberation, and the legitimacy problem for insurers. Philos Public Aff 1997; 4: 303-50.

Dan00 Daniels N. Accountability for reasonableness. Establishing a fair process for priority setting is easier than agreeing on principles. BMJ 2000; 321: 1300-1.

Day98 Day P, Klein R. The dilemmas of choice. Odyssey 1998; 4(2): 8-13.

Dol96 Dolan P. The effect of experience of illness on health state valuations. J Clin Epidemiol 1996; 49: 551-64.

Dol99 Dolan P. Whose preferences count? Med Decis Making 1999; 19: 482-6.

Dol00 Dolan P. The measurement of health-related quality of life for use in resource allocation decisions in health care. In: Culyer AJ, Newhouse JP, editors. Handbook of Health Economics. Volume 1. Amsterdam: Elsevier, 2000: 1724-60.

Don88 Donaldson C, Atkinson A, Bond J, e.a. Should QALYs be programme-specific? J Health Economics 1988; 7: 239-57.

Don97 Donaldson C, Shackley P. Does “proces utility” exist? A case study of willingness to pay for laparoscopic cholecystectomy. Soc Sci Med 1997; 44: 699-707.

Dru93 Drummond M, Torrance G, Mason J. Cost-effectiveness league tables: more harm than good? Soc Sci Med 1993; 37: 33-40.

Dru97 Drummond MF, O’Brien B, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. 2e editie. Oxford: Oxford University Press, 1997.

Edg98 Edgar A, Salek S, Shickle D, e.a. The Ethical Qaly: Ethical issues in healthcare resource allocations. Haslemere: Euromed Communications, 1998.

Enc97 van Enckevort P, Ten Vergert E, Geertsma B. Kosteneffectiviteitsanalyse van longtransplantatie. Medisch Contact 1997; 52: 439-41.

Eme89 Emery DD, Schneiderman LJ. Cost-effectiveness analysis in health care. Hastings Center Report 1989; 19: 8-13.

Ess95 Essink-Bot M L. Health status as a measure of outcome of disease and treatment (Proefschrift). Rotterdam: Erasmus Universiteit, 1995.

Ess00 Essink-Bot ML. Beschrijvend kwaliteit van leven onderzoek. In: Rutten-van Mölken MPMH, van Busschbach JJ, Rutten FFH, editors. Van kosten tot effecten. Een handleiding voor evaluatiestudies in de gezondheidszorg. Maarssen: Elsevier gezondheidszorg, 2000: 63-76.

Eur90 EuroQuol Group. EuroQuol – a new facility for the measurement of health-related quality of life. Health Policy 1990; 16: 199-208.

Fee96 Feeny DH, Torrance GW, Furlong WJ. Health Utilities Index. In: Spilker B, editor. Quality of life and pharmacoeconomics in clinical trials. 2e editie. Philadelphia: Lippincott-Raven, 1996: 239-52.

Fei87 Feinstein AR. Clinimetrics. New Haven: Yale University Press, 1987.

Fis91 Fischoff B. Value elicitation. Is there anything there? American Psychologist 1991; 46: 835-47.

Fow95 Fowler FJ, Cleary PD, Massagli MP, e.a. Describing and measuring the values of health states. The role of reluctance to give up life in the measurement of the values of health states. Med Decis Making 1995; 15: 195-200.

Fri80 Fries JF, Spitz P, Kraines RG, e.a. Measurement of patiënt outcome in arthritis. Arthritis Rheum 1980; 27: 1344-52.

Fro89a Froberg DG, Kane RL. Methodology for measuring health-state preferences. I. Measurement strategies. J Clin Epidemiol 1989; 42: 675-85.

Fro89b Froberg DG, Kane RL. Methodology for measuring health-state preferences. II. Scaling methods. J Clin Epidemiol 1989; 42: 675-85.

Fro89c Froberg DG, Kane RL. Methodology for measuring health-state preferences. III. Population and context effects. J Clin Epidemiol 1989; 42: 675-85.

Fro89d Froberg DG, Kane RL. Methodology for measuring health-state preferences. IV. Progress and research agenda. J Clin Epidemiol 1989; 42: 675-85.

Fry93 Fryback DG, Dasbach EJ, Klein R, e.a. The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors. Med Decis Making 1993; 13: 89-102.

Gab99 Gabriel SE, Kneeland TS, Melton LJ, e.a. Health related quality of life in economic evaluations for osteoporosis: Whose values should we use ? Med Decis Making 1999; 19: 141-8.

Gaf95 Gafni A, Birch S. Preferences for outcomes in economic evaluation: an economic approach to addressing economic problems. Soc Sci Med 1995; 40: 767-76.

Gar96 Garber AM, Weinstein MC, Torrance GW, Kamlet MS. Theoretical foundations of cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996: 25-53.

Gar97 Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. J Health Economis 16, 1997: 1-31.

Gar99 Garber AM. Realistic rigor in cost-effectiveness methods. Med Decis Making 1999; 19: 378-9.

Gar00 Garber AM. Advances in cost-effectiveness analysis of health interventions. In: Culyer AJ, Newhouse JP, editors. Handbook of Health Economics. Volume 1. Amsterdam: Elsevier, 2000: 181-221.

Ger93 Gerard K, Mooney G. QALY league tables: handle with care. Health Econ 1993; 2: 59-64.

Gli99 Glick HA, Polsky D, Willke RJ, e.a. A comparison of preference assessment instruments used in a clinical trial: Responses to the visual analog scale from the EuroQol EQ-5D and the Health Utilities Index. Med Decis Making 1999; 19: 265-75.

Gol96a Gold MR, Patrick DL, Torrance GW, e.a. Identifying and valuing outcomes. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996: 82-134.

Gol96b Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996.

Goo96 Goold SD. Allocating health care: cost-utility analysis, informed democratic decision making, or the veil of ignorance? J Health Politics Policy Law 1996; 21: 69-98.

GR86 Gezondheidsraad. Grenzen van de gezondheidszorg. Den Haag: Gezondheidsraad, 1986; publicatie nr 1986/29.

Gre01 Green C. On the societal value of health care: what do we know about the person trade-off technique? Health Econ 2001; 10: 233-43.

Gre02 Greenberg D, Pliskin JS. Preference-based outcome measures in cost-utility analyses. A 20-year overview. Int J Technol Assess 2002; 18: 461-6.

Gui99 Guillemin F. The value of utility: Assumptions underlying preferences and quality adjusted life years. J Rheumatol 1999; 26: 1861-3.

Guy96 Guyatt GH, Jeaschke R, Feeny DH, e.a. Measurements in clinical trials: choosing the right approach. In:Spilker B, editor. Quality of life and pharmacoeconomics in clinical trials. 2e editie. Philadelphia: Lippincott-Raven, 1996: 41-8.

Had91 Hadorn DC. Setting health care priorities in Oregon. Cost-effectiveness meets the rule of rescue. JAMA 1991; 265(17): 2218-25.

Hal92 Hall J, Gerard K, Salkeld G, e.a. A cost utility analysis of mammography screening in Australia. Soc Sci Med 1992; 34: 993-1004.

Ham97 Ham C. Priority setting in health care: learning from international experience. Health Policy 1997; 42: 49-66.

Har87 Harris J. QALYfying the value of life. J Med Ethics 1987; 13: 117-23.

Har01 Hart W, Walvoort HC. ‘Troosten: altijd’; condoleancebetuigingen van artsen aan de nabestaanden van hun patiënten. Ned Tijdschr Geneeskd 2001; 145: 2185-7.

Hey90 Heyink J. Adding years to your life or adding life to your years. QALY (Quality Adjusted Life Year): theoretical and practical issues. Int J Health Sciences 1990; 1: 45-9.

Hol95 Hollingworth W, Mackenzie R, Todd CJ, e.a. Measuring changes in quality of life following magnetic resonance imaging of the knee: SF-36, EuroQol or Rosser index? Qual Life Res 1995; 4: 325-34.

Hol98 Holm S. Goodbye to the simple solutions: the second phase of priority setting in health care. BMJ 1998; 317: 1000-2.

Hou98 van Hout BA. Discounting costs and effects: a reconsideration. Health Econ 1998; 7: 581-594.

Hou01 van Hout BA. Geen Sinterklaas zonder Zwarte Piet. Over keuzen in de gezondheidszorg (Inaugurele rede). Utrecht: Universiteit Utrecht, 2001.

Hun80 Hunt SM, McEwen J. The development of a subjective health indicator. Sociol Health Illness 1980; 2: 231-46.

Hun98a Hunt SM. Cross-culturual issues in the use of quality of life measures in randomized controlled trials. In: Staquet MJ, Hays RD, Fayers PM, editors. Quality of life assessment in clinical trials. Oxford: Oxford University Press, 1998: 51-67.

Hun98b Hunink MGM, Bult JR, de Vries J, e.a. Uncertainty in decision models analyzing cost-effectiveness. Med Decis Making 1998; 18: 337-46.

Hur98 Hurley J. Welfarism, extra-welfarism and evaluative economic analysis in the health sector. In: Barer ML, Getzen TE, Stoddart GL, editors. Health, health care and health economic: perspectives on distribution. Chicester: Wiley, 1998: 373-95.

Hur00 Hurley J. An overview of the normative economics of the health sector. In: Culyer AJ, Newhouse JP, editors. Handbook of health economics. Vol 1. Amsterdam: Elsevier, 2000: 56-118.

Hut96 Hutchinson JF. Quality of life in ethnic groups. In: Spilker B, red. Quality of life and pharmacoeconomics in clinical trials. 2e editie. Philadelphia: Lippincott-Raven, 1996: 587-93.

Joh96 Johannesson J, Jönsson B, Karlsson G. Outcome measurement in economic evaluation. Health Econ, 1996; 5: 279-96.

Joh97 Johannesson M, Meltzer D, O'Conor RM. Incorporating future costs in medical cost-effectiveness analysis: implications for the cost-effectiveness of the treatment of hypertension. Med Decis Making 1997; 17: 382-9.

Jon86 Jonsen AR. Bentham in a box: technology assessment and health care allocation. Law Med Health Care 1986; 14(3-4): 172-4.

Kah79 Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica 1979; 47: 263-91.

Kar48 Karnofsky DA, Abelman WH, Craver LF, e.a. The use of nitrogen mustards in the palliative treatment of carcinoma. Cancer 1948; I: 634-56.

Kar49 Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In: MacLeod CM, red. Evaluation of chemotherapeutic agents. New York: Columbia University Press, 1949: 191-205.

Kat63 Katz S, Ford AB, Moskowitz RW, e.a. Studies of illness in the aged. The index of ADL. JAMA 1963; 185: 914-9.

Kee83 Keeler EB, Cretin S. Discounting of life-saving and other nonmonetary effects. Management Science 1983; 29: 300-6.

Keu91 Commissie Keuzen in de zorg. Kiezen en delen. 1. Den Haag: DOP, 1991.

Kin96 Kind P. The EuroQol instrument: an index of health-related quality of life. In: Spilker B, red. Quality of life and pharmacoeconomics in clinical trials. 2e editie. Philadelphia: Lippincott-Raven, 1996: 191-201.

Kir00 Kirsch J, McGuire A. Establishing health state valuations for disease specific states: an example from heart disease. Health Economic 2000; 9: 149-58.

Kle93 Klein R. Health care reform: the global search for Utopia. Br Med J 1993; 307: 752.

Klei98 Klein R. Puzzling out priorities. Why we must acknowledge that rationing is a political process. Br Med J 1998; 317: 959-60.

Kon91 de Koning HJ, van Ineveld BM, van Oortmarssen GJ, e.a. Breast cancer screening and cost-effectiveness: Policy alternatives, quality of life considerations and the possible impact of uncertain factors. Int J Cancer 1991; 49: 531-7.

Koo94 Koopmanschap MA, Rutten FFH. The impact of indirect costs on outcomes of health care programs. Health Econ 1994; 3: 385-93.

Koo95 Koopmanschap MA, Rutten FFH, Ineveld BM, e.a. The friction cost method for measuring indirect costs of disease. J Health Economics 1995; 14: 171-89.

Koo98 Koopmanschap MA, Rutten FFH. Berekening van kosten van zorg. Vaak onderschat in economische evaluatiestudies. Tijdschr Gezondheidswetens 1998; 76: 83-8.

Kor99 Korczyn AD, Davidson M. Quality of life in Alzheimer's disease. Eur J Neurol 1999; 6: 487-9.

Kra93 Krahn M, Gafni A. Discounting in the economic evaluation of health care interventions. Med Care 1993; 31: 403-18.

Kra98a Krabbe PFM, Bonsel GJ. Sequence effects, health profiles, and the QALY model: in search of realistic modelling. Med Decis Making 1998; 18: 178-86.

Kra98b Krabbe PFM. The valuation of health outcomes. A contribution to the QALY approach (Proefschrift). Rotterdam: Erasmus Universiteit, 1998.

Leg02 Legemaate J. De (dubbel)rol van de arts. In: Zorg, schaarste en recht. Preadvies uitgebracht ten behoeve van het Jubilieumsymposium van de Vereniging voor Gezondheidsrecht. Vereniging voor Gezondheidsrecht, 23 mei 2002: 34-63.

Len99 Lenert LA, Treadwell JR, Schwartz CE. Associations between health status and utilities. Implications for policy. Med Care 1999; 37: 479-89.

Len00 Lenert L, Kaplan RM. Validity and Interpretation of preference-based measures of health-related quality of life. Med Care 2000; 38 (suppl II): II-138-145.

Lep97 Leplège A, Hunt S. The problem of quality of life in medicine. JAMA 1997; 278: 47-50.

Leu93 Leussink AB. Kiezen in de zorg: het meten van het onmeetbare. QALY of SAVE. Graadmeter 1993; 9: 5-7.

Lle02 Llewellyn-Thomas HA, Arshinoff R, Bell M, e.a. Healthy year equivalents in major joint replacement. Can patients provide meaningful responses? Int J Technol Assess Health Care 2002; 18: 467-84.

Lip96 Lipscomb J, Weinstein MC, Torrance GW. Time preference. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996: 214-46.

Loo89 Loomes G, McKenzie L. The use of QALYs in health care decision making. Soc Sci Med 1989; 28: 299-308.

Luc96 Luce BR, Manning WG, Siegel JE, e.a. Estimating costs in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996: 176-213.

Maa95 van der Maas PJ. Volksgezondheidsonderzoek. In: van der Maas PJ, Mackenbach JP, editors. Volksgezondheid en gezondheidszorg. Utrecht: Bunge, 1995: 17-75.

Mac01 Macran S, Kind P. "Death" and the valuation of health-related quality of life. Med Care 2001; 39: 217-27.

Man96a Mandelblatt JS, Fryback DG, Weinstein MC, e.a. Assessing the effectiveness of health interventions. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996: 135-75.

Man96b Manning WG, Fryback DG, Weinstein MC. Reflecting uncertainty in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996: 247-75.

McK98 McKie J, Richardson J, Singer P, e.a. The allocation of health care resources. An ethical evaluation of the 'Qaly' approach. Aldershot: Ashgate/Dartmouth, 1998.

Meh89 Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. Med Decis Making 1989; 9: 142-9.

Mel75 Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. Pain 1975; 1: 277-99.

Mel97 Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. J Health Economics 1997; 16: 33-64.

Mel99a Meltzer D, Johannesson M. Inconsistencies in the "societal perspective" on costs of the Panel on cost-effectiveness in health and medicine. Med Decis Making 1999; 19: 371-7.

Mel99b Meltzer D, Johannesson M. On the role of theory in cost-effectiveness analysis – a response to Garber, Russell, and Weinstein. Med Decis Making 1999; 19: 383-4.

Mel00 Meltzer D, Egleston B, Stopffel D, e.a. Effect of future costs on cost-effectiveness of medical interventions among young adults. The example of intensive therapy for type 1 diabetes mellitus. Medical Care 2000; 38: 679-85.

Men90 Menzel PT. Strong medicine. The ethical rationing of health care. New York, Oxford: Oxford University Press, 1990: 79-96.

Men99 Menzel P, Gold MR, Nord E, e.a. Toward a broader view of values in cost-effectiveness analysis of health. Hastings Center Report 1999; May-June: 7-15.

Men02 Menzel P, Dolan P, Richardson J, e.a. The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. Soc Sci Med 2002; 55: 2149-58.

Miy85 Miyamoto JM, Eraker SA. Parameter estimates for a QALY utility model. Med Decis Making 1985; 5: 191-213.

Moo94a Mooney G. What else doe we want from our health services? Soc Sci Med 1994; 39: 151-4.

Moo94b Mooney G. Key issues in health economics. New York: Harvester Wheatsheaf, 1994.

Moo00 Mooney G, Wiseman V. Burden of disease and priority setting. Health Econ 2000; 9: 369-72.

Mul97 Mulder JH. Terug naar de zorg. Dunning gekanteld. Med Contact 1997; 52: 1219-20.

Mur96 Murray CJL, Lopez AD. The global burden of disease. A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Cambridge, MA: Harvard University Press, 1996.

Mur97 Murray CJL, Acharya AK. Understanding DALYs. J Health Economics 1997; 16: 703-30.

Mus87 Musschenga AW, Borst-Eilers E. Ethische criteria voor prioriteitenstelling in de gezondheidszorg. In: Musschenga AW, de Neeling JND, editors. Verdeling van schaarse middelen in de gezondheidszorg. Amsterdam: VU Uitgeverij, 1987: 101-22.

Nea95 Nease RF, Kneeland T, O'Connor GT, e.a. Variation in patient utilities for outcomes of the management of chronic stable angina: implications for clinical practice guidelines. JAMA 1995; 272: 1185-90.

Nee91 de Neeling JND. Quality of life. Het onderzoek naar welzijnseffecten van medische behandelingen. Utrecht: Bunge, 1991.

Nee92 de Neeling JND. Kwaliteit van leven: een meetbaar begrip? Medisch Contact 1992; 47: 143-5.

Nee00 de Neeling JND, namens de projectgroep Passende Medische Zorg. Passende medische zorg. Over keuzen en richtlijnen. Eindrapport KNMG-project Passende Medische Zorg. Utrecht: KNMG, 2000.

Neu94 Neumann PJ, Johannesson M. The willingness to pay for in vitro fertilization: a pilot study using contingent valuation. Med Care 1994; 32: 686-99.

Neu97 Neumann PJ, Zinner DE, Wright JC. Are methods for estimating QALYs in cost-effectiveness analyses improving? Med Decis Making 1997; 17: 402-8.

Neu99 Neumann PJ, Hermann RC, Kuntz KM, e.a. Cost-effectiveness of donepezil in the treatment of mild or moderate Alzheimer's disease. Neurology 1999; 52: 1138-45.

Neu00a Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. Annu Rev Public Health 2000; 21: 587-611.

Neu00b Neumann PJ, Sandberg EA, Araki SS, e.a. A comparison of HUI2 and HUI3 utility scores in Alzheimer's disease. Med Decis Making 2000; 20: 413-22.

Nor92a Nord E. An alternative to QALYs: the saved young life equivalent (SAVE). BMJ 1992; 305: 875-7.

Nor92b Nord E. Methods for quality adjustment of life years. Soc Sci Med 1992; 34: 559-69.

Nor93 Nord E. The trade-off between severity of illness and treatment effect in cost-value analysis of health care. Health Policy 1993; 24: 227-38.

Nor94 Nord E. The QALY – a measure of social value rather than individual utility? Health Economics 1994; 3: 89-93.

Nor95 Nord E. The person-trade-off approach to valuing health care programs. Med Decis Making 1995; 15: 201-8.

Nor99 Nord E, Pinto JL, Richardson J, e.a. Incorporating societal concerns for fairness in numerical valuations of health programmes. Health Econ 1999; 8: 25-39.

Nor01 Nord E. The desirability of a condition versus the well being and worth of a person. Health Econ 2001; 10: 579-81.

Oos00 Oostenbrink JB, Koopmanschap MA, Rutten FFH. Handleiding voor kostenonderzoek. Methoden en richtlijnprijzen voor economische evaluaties in de gezondheidszorg. Amstelveen: College voor Zorgverzekeringen, 2000.

Pat89 Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Med Care 1989; 27 suppl: S217-32.

Pin97 Pinto Prades JL. Is the person trade-off a valid method for allocating health care resources? Health Econ 1997; 6: 71-81.

Pli80 Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. Opns Res 1980; 28: 206-24.

Pol02 Poley MJ, Stolk EA, Brouwer WBF, van Busschbach JJ. Ziektelast als uitwerking van het criterium 'noodzakelijkheid' bij het maken van keuzen in de zorg. Ned Tijdschr Geneeskd 2002; 146: 2312-5.

Pro00 Prosser LA, Stinnett AA, Goldman PA, e.a. Cost-effectiveness of cholesterol-lowering therapies according to selected patient characteristics. Ann Int Med 2000; 132: 769-79.

Raw89 Rawles J. Castigating QALYs. J Med Ethics 1989; 15: 143-7.

Rea84 Read JL, Quinn RJ, Berwick DM, e.a. Preferences for health outcomes: comparisons of assessment methods. Med Decis Making 1984; 4: 315-29.

RGO01 Raad voor Gezondheidsonderzoek. Rapport van het Platform HTA. Bijlage 1. Rapport van de commissie Methodologie. Den Haag: Raad voor Gezondheidsonderzoek, 2001.

Ric94 Richardson J. Cost utility analysis: what should be measured? Soc Sci Med 1994; 39: 7-21.

Ric96 Richardson J, Hall J, Salkeld G. The measurement of utility in multiphase health states. Int J Technol Assess Health Care 1996; 12: 151-62.

Rie98 Ried W. QALYs versus HYEs – what's right and what's wrong. A review of the controversy. J Health Econ 1998; 17: 607-25.

Rig86 Rigter H. Wie volgt? Over selectie en wachtlijsten in de gezondheidszorg. Den Haag: Gezondheidsraad 1986; publicatie nr 1986/28.

Rus96 Russell LB, Siegel JE, Daniels N, e.a. Cost-effectiveness analysis as a guide to resource allocation in health: roles and limitations. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996: 3-24.

Rus99 Russell LB. Improving the panel's recommendations. Med Decis Making 1999; 19: 379-80. (Archief: 229)

Rut98 Rutten FFH, van Busschbach JJ, van Hout BA, e.a. Economische evaluatie van gezondheidszorgprogramma's. Principes en instrumentarium. Tijdschr Gezondheidswetens 1998; 76: 74-82.

Rut00a Rutten FFH. Voor de praktijk. Dwalingen in de methodologie. XIX. Kosteneffectiviteitsratio's. Ned Tijdschr Geneekd 2000; 144: 1311-14.

Rut00b Rutten FFH, Oostenbrink JB, Koopmanschap MA. Directe kosten van gezondheidszorg. In: Rutten-van Mölken MPMH, Van Busschbach JJ, Rutten FFH, editors. Van kosten tot effecten. Een handleiding voor evaluatiestudies in de gezondheidszorg. Maarssen: Elsevier gezondheidszorg, 2000: 31-41.

Sac78 Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. J Chron Dis 1978; 31: 697-704.

Sas01 Sassi F, Archard L, Le-Grand J. Equity and the economic evaluation of healthcare. Health Technol Assessment 2001; 5(3): 1-138.

Sen91 Sen A. Utility. Ideas and terminology. Economics and Philosophy 1991; 7: 277-83.

Shi97 Shiell A. Health outcomes are about choices and values: an economic perspective on the health outcomes movement. Health Policy 1997; 39: 5-15.

Slo95 Slovic P. The construction of preference. American Psychologist 1995; 50: 364-71.

Smi87 Smith A. Qualms about QALY's. Lancet 1987; May 16: 1134-6.

Smi93 Smith R, Dobson M. Measuring utility values for QALYs: two methodological issues. Health Econ, 1993; 2: 349-55.

Spi90 Spilker B, Molinek FR, Johnston KA, e.a. Quality of life bibliography and indexes. Med Care 1990; 28 suppl: DS1-77.

Spi96 Spilker B, red. Quality of life and pharmacoeconomics in clinical trials. 2e editie. Philadelphia: Lippincott-Raven, 1996.

Ste00 Stewart AL, Napoles-Springer A. Health-related quality-of-life assessments in diverse population groups in the United States. Med Care 2000; 38, Suppl: II-102-24.

Sto00a Stouthard MEA, Essink-Bot ML, Bonsel GJ, e.a. Disability weights for diseases. A modified protocol and results for a Western European region. Eur J Public Health 2000; 10: 24-30.

Sto00b Stone PW, Chapman RH, Sandberg EA, e.a. Measuring costs in cost-utility analyses. Variations in the literature. Int J Technol Assessment Health Care 2000; 16: 111-24.

Sug78 Sugden R, Williams A. The principles of practical cost-benefit analysis. Oxford: Oxford University Press, 1978.

Tor86 Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ 1986; 5: 1-30.

Tor89 Torrance GW, Feeny D. Utilities and quality-adjusted life years. Int J Technology Assessment Health Care, 1989; 5: 559-75.

Tor99 Torgersen D, Raftery J. Measuring outcomes in economic evaluations. Br Med J 1999; 318: 1413.

Tor86 Torrance GW. Measurement of health state utilities for economic appraisal. J Health Economics 1986; 5: 1-30.

Tsu00 Tsuchiya A. QALYs and ageism: philosophical theories and age weighting. Health Econ 2000; 9: 57-68.

Ube99 Ubel PA, Richardson J, Prades J LP. Life-saving treatments and disabilities. Are all QALYs created equal? Int J Technol Assessment Health Care 1999; 15: 738-48.

Ube00a Ubel PA. Pricing life. Why it's time for health care rationing. Cambridge, MA: MIT Press, 2000.

Ube00b Ubel PA, Richardson J, Menzel P. Societal value, the person trade-off, and the dilemma of whose values to measure for cost-effectivenes analysis. Health Econ 2000; 9: 127-36.

Ube00c Ubel PA, Nord E, Gold M, e.a. Improving value measurement in cost-effectiveness analysis. Med Care 2000; 38(9): 892-901.

Uyl00 Uyl-de Groot CA. Uitgangspunten en opzet van onderzoek. In: Rutten-van Mölken MPMH, van Busschbach JJ, Rutten FFH, editors. Van kosten tot effecten. Een handleiding voor evaluatiestudies in de gezondheidszorg. Maarssen: Elsevier gezondheidszorg, 2000: 17-30.

Ver01 Verkerk MA, van Busschbach JJ, Karssing ED. Health-related quality of life research and the capability approach of Amartya Sen. Qual Life Res 2001; 10: 49-55.

Vle90 Vlek CAJ. Beslissen over risico-acceptatie. Den Haag: Gezondheidsraad, 1990; publicatie nr A90/10.

Wag00 Wagemakers A, Van der Wijk P, Francissen P (red). Wanneer is het leven te kostbaar? Leeftijd en kosteneffectiviteit in de gezondheidszorg. Amsterdam: Boom, 2000.

Wag91 Wagstaff A. QALYs and the equity-efficiency trade-off. J Health Econ 1991; 10: 21-41.

Wei80 Weinstein MC, Fineberg HV, editors. Clinical decision analysis. Philadephia: WB Saunders, 1980.

Wei88 Weinstein MC. A QALY is a QALY is a QALY – or is it? J Health Economics 1988; 7: 289-90.

Wei96 Weinstein MC, Siegel JE, Gold MR, e.a. Recommendations of the panel on cost-effectiveness in health and medicine. JAMA 1996; 276: 1253-8.

Wei97a Weinstein MC, Siegel JE, Garber AM, e.a. Productivity costs, time costs and health-related quality of life: a response to the Erasmus group. Health Econ 1997; 6: 505-10.

Wei97b Weinstein MC, Manning WG. Theoretical issues in cost-effectiveness analysis. J Health Economics 1997; 16: 121-8.

Wei99 Weinstein MC. Theoretically correct cost-effectiveness analysis. Med Decis Making 1999; 19: 381-2.

Wil81 Williams A. Welfare economics and health status measurement. In: Gaag J van der, Perlman M, editors. Health, economics and health economics. Amsterdam: North Holland, 1981.

Wil92 Williams A. Cost-effectiveness analysis: is it ethical? J Med Ethics 1992; 18: 7-11.

Wil93 Williams A. Cost-benefit analysis: applied welfare economics or general decision aid. In: Williams A, Giardina E, editors. Efficiency in the public sector. London: Edward Elgar, 1993.

Wil96 Williams A. QALYs and ethics: a health economist's perspective. Soc Sci Med 1996; 43: 1795-1804.

Wil00 Williams A, Cookson R. Equity in health. In: Culyer AJ, Newhouse JP, editors. Handbook of Health Economics. Volume 1. Amsterdam: Elsevier, 2000: 1863-1910.

Wit00 de Wit GA, van Busschbach JJ, de Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? Health Econ 2000; 9: 109-26.

Wor93 World Bank. World development report 1993: investing in health. New York: Oxford University Press, 1993.

WRR97 Wetenschappelijke Raad voor het Regeringsbeleid. Volksgezondheidszorg. Den Haag: Sdu, 1997.

Zec76 Zeckhauser R, Shepard D. Where now for saving lives? Law and Contemporary Problems 1976; 40: 5-45.

A Responsibility

Annex

Annex **A**

Responsibility

This report was discussed in draft form in the following Health Council committees:

- Committee on Contours of the Basic Health Care Benefit Package
- Core Committee on Medical Technology Assessment
- Standing Committee on Medicine
- Standing Committee on Medical Ethics and Health Law

Furthermore, the following external experts have commented on a draft of this report:

- H Bleichrodt, PhD, economist; Erasmus University Rotterdam
- JJ van Busschbach, PhD, psychologist; Erasmus University Rotterdam
- Professor BA van Hout, PhD, economist; University Medical Centre Utrecht
- Professor FFH Rutten, PhD, economist; Erasmus University Rotterdam

Responsibility for the text rests entirely with the author, who is the secretary to the Health Council's Standing Committee on Medicine.