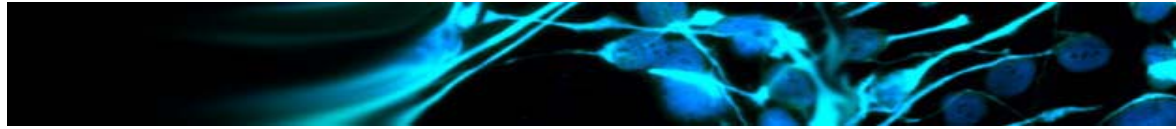


# *Occasional Paper 11/03*

## Measuring Value: Pharmacoeconomics in Theory and Practice

Adrian Towse, OHE  
Michael Drummond, University of York  
Corinna Sorenson, LSE

21 May 2011



# Measuring Value: Pharmacoeconomics in Theory and Practice

Office of Health Economics  
Occasional Paper 11/03  
21 May 2011



Adrian Towse, OHE  
Michael Drummond, University of York  
Corinna Sorenson, LSE

Corresponding author: Adrian Towse

<sup>1</sup> Southside, 7th Floor, 105 Victoria Street, London SW1E 6QT Tel: 020 7747 8855  
E-mail: [atowse@ohe.org](mailto:atowse@ohe.org)

### **About the authors**

Adrian Towse is Director of the Office of Health Economics and Visiting Professor at the Department of Economics and Related Studies at the University of York.

Mike Drummond is Professor of Health Economics at the University of York and Part-time Professor at LSE Health at the London School of Economics and Political Science.

Corinna Sorenson is a Research Officer in European Health Policy and PhD candidate at LSE Health at the London School of Economics and Political Science.

### **Acknowledgement and Disclaimer**

A version of this Occasional paper will be published as a chapter in the Oxford Handbook on the Economics of the Biopharmaceutical Industry, edited by Danzon and Nicholson, to be published in 2012.

This Occasional Paper is the work of the authors and does not necessarily represent the views of the Office of Health Economics, the University of York, or of the London School of Economics and Political Science. All errors remain the responsibility of the authors.

## Contents

1. Introduction .....	5
2. The evolution of CEA from the practice of Health Technology Assessment (HTA).....	7
3. The theoretical and practical issues that have arisen in the application of CEA to the regulation of pharmaceutical prices and use. ....	10
4. Valuing incremental health gain .....	12
4.1 Strengths and Weaknesses of the QALY.....	12
4.1.1. Origins and Definition of the QALY .....	12
4.1.2 Valuing health states – which instrument to use?.....	13
4.1.3. Valuing health states – whose values to use?.....	13
4.1.4. Valuing health states – profiling over time .....	15
4.1.5. Valuing health states – concerns for fairness.....	15
4.2. Extra-welfarism versus welfarist use of the QALY .....	17
4.3 Alternatives to the QALY .....	18
4.4 The use of WTP methods to value health care interventions .....	19
5. Which costs and benefits to include alongside health gain? .....	21
6. What types of evidence to use?.....	23
6.1 Economic evaluation in drug development.....	23
6.2 Issues in the transferability of economic evaluations .....	25
6.3 Towards an International Standard for Economic Evaluation.....	26
7. Making decisions.....	28
7.1 Weighting different factors in a deliberative process .....	28
7.2 Using decision support tools such as MCDA.....	29
7.3 The opportunity cost of adopting a technology .....	29
7.4 Implementation of cost-effective treatments.....	32
7.5 Disinvestment from treatments that are not cost-effective .....	33
8. Handling evidence uncertainty in decision making.....	35
8.1 The analysis and reporting of uncertainty with sensitivity analysis.....	35
8.2 Value of information and coverage with evidence development.....	35
9. Are decision makers getting what they want in an efficient manner? .....	38
9.1 Setting priorities for undertaking studies.....	38
9.2 The costs of studies need to reflect the expected benefits .....	38
9.3 Decision maker ability to interpret and use studies.....	39
10. Conclusions .....	40
References .....	42



## 1. Introduction

Government-run or regulated health care systems have adopted complex systems of economic regulation to control pharmaceutical expenditures, which are generally an ad hoc mix of historical policies. There has been a substantial increase in the number of third party payers using formal “cost-effectiveness analysis” (CEA), however, or pharmacoeconomic approaches for assessing the value of drugs, vaccines, and other health technologies to inform decisions about pricing, reimbursement, and use within their health care systems.

CEA is usually applied within a broader framework of health technology assessment (HTA). With this approach, drugs and other technologies are assessed for use and/or for a reimbursement price by looking at incremental health-related effects and costs relative to existing treatments. Australia was the first jurisdiction to adopt such a policy in 1993 and was quickly followed by New Zealand and several Canadian provinces. The United Kingdom (UK) established the National Institute for Health and Clinical Excellence (NICE) in 1999 to review the efficacy and cost of technologies expected to have major health or budgetary impact, using cost per quality adjusted life year (QALY), and to formulate guidance on the use of these technologies in the National Health Service (NHS) in England and Wales. In Sweden, the Dental and Pharmaceutical Benefits Board (TLV) undertakes CEA to inform decisions on the reimbursement of drugs. Other European countries requesting economic submissions for some, or all, new medicines, include Belgium, Finland, Ireland, Norway, The Netherlands, Portugal and Germany<sup>1</sup>. Similar policies have also been recently adopted by some countries in Eastern European, Asia (e.g., South Korea) and Latin America (e.g., Brazil).

Within the United States (US), HTA and CEA are much less established. However, over the last couple of years, there has been growing interest in establishing a more formalized process or system for conducting comparative-effectiveness research (CER)<sup>2</sup>. The 2010 US Patient Protection and Affordable Care Act (PPACA) sets up a Patient Centered Outcomes Research Institute (PCORI) tasked to commission CER. However, it specifically forbids the use of “a dollars per quality adjusted life year (or similar measure that discounts the value of a life year because of an individual’s disability) as a threshold to establish what type of health care is cost effective or recommended..” (PPACA, 2010). This reflects political concern about the ethics of the QALY

---

<sup>1</sup> Germany established the Institute for Quality and Efficiency in Health Care (IQWiG) in 2004 to evaluate technologies for the Joint Federal Committee (G-BA). IQWiG’s role was expanded to include CEA of drugs, although there were substantial delays in agreeing a methodology. 2010 legislation requires all companies seeking to avoid a reference price to submit a “value dossier” to the G-BA/IQWiG at the time of product launch.

<sup>2</sup> Luce et al. (2010) explore the relationship between Evidence Based Medicine (EBM), CER and HTA. There is ambiguity as to whether CER includes resource use and costs, (the US PPACA does not include costs or resource use in its definition). It clearly includes evidence on comparative health effects (positive and negative) and as such is an input to an HTA, which is a form of evidence synthesis for decision making. We take CEA to be a form of HTA in which evidence on the incremental costs and effects of a health intervention is analysed. It should therefore include the results of any CER. Pharmacoeconomics is, in effect, CEA for drugs. It is important to separate the analysis or assessment of costs and effects from decision making as to whether or not to use/reimburse a drug at a particular price. The latter requires an *interpretation* of the assessment and *judgement* as to the value of the drug in the light of the evidence.

and about Federal government rationing of health care<sup>3</sup>. There is concern that this could reduce the likelihood of CEA being commissioned by PCORI (Neumann and Weinstein, 2010). Garber and Sox (2010) argue that PCORI commissioned CER should include resource use and so enable others in the public and private sectors to undertake CEA.

The wording of the PPACA means that the generation and use of pharmacoeconomics in the US is likely to remain largely decentralized: each of the 50 state Medicaid programs has some form of HTA procedure for drugs; 13 states participate in the Drug Effectiveness Review Project (DERP) which commissions HTA from a network of academic centres. Many private health plans and pharmacy benefit managers (PBMs) also operate HTA programs, and pharmacoeconomic analyses are frequently conducted by manufacturers, consulting firms, and academic departments.

In this paper we explore the theoretical and practical issues that have arisen in the application of CEA for drugs to resource allocation decisions in health care and in the regulation of pharmaceutical prices and use. We begin by outlining its evolution from the practice of Health Technology Assessment (HTA).

---

<sup>3</sup> During the US 2009/10 debate about health care reform, the UK NICE was frequently referred to as a “death panel” for its role in rationing NHS care.

## 2. The evolution of CEA from the practice of Health Technology Assessment (HTA)

HTA originated from growing concern about the extensive diffusion of costly medical equipment in the 1970s and taxpayers and health insurers' ability and willingness to fund their use (Jönsson and Banta, 1999). Moreover, greater public awareness of health care rationing decisions and a growing consumerist position on health care policy required more accountability, transparency and legitimacy in decision-making processes. Consequently, decision makers needed a more comprehensive approach to set priorities and obtain maximum benefit from limited resources, without compromising the ethical and social values underpinning health systems (Hutton et al., 2006). The growth and development of HTA reflected this demand for information to support decisions on the development, uptake and diffusion of health technologies.

Since the 1970s, HTA has broadened to encompass a wide range of health technologies including high cost surgical procedures, drugs, medical devices, and the organizational and support systems for care provision (Jönsson, 2002). The majority of HTAs are now conducted on pharmaceuticals (Hutton et al., 2006).

The evolution of HTA was mirrored by the development of economic techniques to assess the value of health technologies. Within economics, the standard approach for assessing the efficiency of projects was a cost-benefit analysis (CBA) (Mishan, 1971). In this approach, with its theoretical foundations in welfare economics, monetary valuations of the benefits of projects are obtained through contingent valuation (i.e., assessments of individuals' willingness to pay) and compared with the costs. If the total benefits from the project exceed the total costs, then the project should go ahead<sup>4</sup>. CBA is well-established in the area of public policy, particularly in the fields of environmental economics and transport economics. See Arrow et al. (1993) for practical applications of CBA in environmental economics.

However, in the field of health economics, several concerns about the application of CBA led health economists to develop CEA, a variant of CBA, as an alternative approach (Klarman et al., 1968; Gold et al., 1996; McGuire., 2001). Some of these concerns relate to issues of principle, in that the application of CBA may lead to inequity in access to health care if individuals' willingness to pay is constrained by their ability to pay or that hypothetical "ability to compensate" assessments do not actually lead to equitable outcomes. Others relate to the difficulties of obtaining reliable estimates of the value individuals place on items such as the value of life using willingness-to-pay techniques.

In the CEA approach the analyst does not explicitly assign a monetary value to the health outcomes, as the results are expressed in terms of the incremental cost per unit of health gain. The most widely used unit of measure is the quality-adjusted life-year (QALY), which combines

---

<sup>4</sup> Albeit with the potential to use a shadow price for funds or other approaches to rationing if not all projects with a positive NPV can be funded.



gains in length and “quality” of life. Health states are assigned different levels of “quality”. However, in order to make a judgement of whether an individual project is worthwhile, the incremental cost-effectiveness ratio needs to be compared with an external reference standard or monetary threshold. This external reference standard has to be specified by a decision maker acting as the agent of the third party payer, who in many cases is a public payer acting on behalf of society (Sugden and Williams, 1979), and who may look for comparisons with existing, previously-funded programs (Maynard, 1991), albeit with the need to take account of the overall budget (Birch and Gafni, 2006). Some commentators refer to analytic approaches based on CEA as being “extra-welfarist”, in that they represent an extension of standard welfare economics, by incorporating other arguments, such as health and equity, into the social welfare function (Culyer, 1989). Others (Garber and Phelps, 1997) have offered a framework for grounding CEA and the QALY in welfare economics. We return to these issues later in the paper.

Whilst our focus here is on CEA, HTA is in principle broader. It has been defined as “the systematic evaluation of the properties, effects, and/or other impacts of health care technology” and as including “the medical, economic, social and ethical implications of development, diffusion and use of health technology” (INAHTA, 2002). However, most HTA is now CEA for drugs, and CEA is the dominant approach in pharmacoeconomics. The principal question being addressed in HTAs is “Is the technology worth it?”, which comprises five elements:

- Is the technology effective? What benefits does it bring?
- For whom does the technology work?
- What costs are entailed in its use?
- How does the technology compare with available treatment alternatives?
- Given the answers to these questions, is the technology worth using in the health care system for some or all of those who would benefit?

HTAs are currently being performed by a variety of public and private sector organizations, advisory committees, and regulatory bodies in many (and an increasing number of) jurisdictions. Historically, most HTA agencies have focused on producing high quality assessment reports that can be used by a range of decision makers for a variety of uses. However, as we noted above, use of HTA is now increasingly undertaken within countries to inform resource allocation decisions about the reimbursement, coverage, or use of a health technology, particularly in relation to pharmaceuticals. For a discussion of principles of good practice for use of HTA in resource allocation, see Drummond et al., 2008.

The primary rationale for the use of CEA is therefore to improve resource allocation in health care. It is designed to help any decision maker maximize the value gained from limited resources, in a context where reliance on individual consumer choice to make these decisions is not possible or deemed appropriate. This is the norm, as most health care provision in developed countries is financed via public or private third party payers. Even in a system with

private insurance such as the US, a payer has to make decisions about what services will be covered under what circumstances and how much the health plan will reimburse, and price premiums accordingly<sup>5</sup>. CEA offers a framework for making those decisions.

CEA for drugs also has the effect of regulating drug prices indirectly through a review of cost-effectiveness. This is in theory more consistent with principles of efficient resource allocation than other regulatory methods for drug prices. It means that more effective/safer drugs (delivering more QALYs) can charge higher prices and still be cost-effective relative to less effective/less safe drugs. This provides efficient incentives for research and development (R&D). In addition, by using the CEA approach, the indications in which the use of the drug would be efficient can also be identified and potentially controlled, thereby encouraging a more cost-effective use of the drug in the health care system. This is far more efficient than, for example, controlling expenditure through drug budget caps, a form of “silo budgeting” (Garrison and Towse, 2003) which creates perverse incentives for cost shifting to less efficient inputs or to curb sales of products that are delivering a lot of health gain, or by simply ignoring potential differences between drugs in a therapeutic group and pricing them all at the same price (Drummond et al., 2010).

Improved static and dynamic efficiency was behind the UK OFT proposal for a system of VBP for drugs (Office of Fair Trading, 2007), whereby prices are based on incremental health gain net of incremental cost. This led to changes in UK price regulation (Department of Health, 2009; 2010). VBP raises issues as to how health gain is measured and valued and what other costs and effects are taken into account (Claxton, 2007; Towse, 2007). It also introduces a number of important “process” questions, notably whether prices are set by governments based on “value” or by companies, with governments and other payers deciding whether to use and/or reimburse the product based on the price and evidence of its “value.” For an early discussion of these issues see Drummond et al. (1997).

---

<sup>5</sup> We discussed earlier the sensitivity to the use of CEA and QALYs in the US public sector. Use of CEA is also a sensitive matter in the US private sector (Neumann, 2004). The US norm is that public and private health plans pay for “necessary and reasonable” care. Leaving this decision to physicians and patients may result in overuse of services, if moral hazard is present, and there are no other constraints. Reasonableness, however, must have regard to the resources available and the efficiency of the treatment. CEA provides information on these.

### **3. The theoretical and practical issues that have arisen in the application of CEA to the regulation of pharmaceutical prices and use.**

The formal use of CEA by bodies setting reimbursement prices or utilization guidelines for drugs has given rise to a series of practical issues. Addressing a number of these issues has stimulated developments in theory and methods. The main issues, addressed in the remainder of this paper, are as follows:

#### **1. *How should the incremental health gain that a technology delivers be valued?***

The strengths and weaknesses of the QALY, the measure of health gain most used in practice, are discussed as well as the challenges of using other approaches, such as WTP to measure value.

#### **2. *What else should be included in assessments besides health gain to the patient and costs to the health care system?***

The debate as to whether to take into account the benefits of getting people back to work, the impact on carers, or changes in other non-health effects arising from the use of an intervention is reviewed.

#### **3. *What types of evidence matter?***

A number of issues arise in considering what types of evidence to require or include in assessments. This section discusses the handling of possible heterogeneity of response within the patient population; translating evidence from one jurisdiction to another; the potential for greater coordination of licensing and HTA data requirements and the implications for global development programmes; collecting data alongside use of the product; and, synthesising evidence from different studies.

#### **4. *How should the HTA decision-making body make and implement decisions?***

Four issues are covered: 1) how bodies should weigh the different factors together in their deliberative processes; 2) whether there is a case for using decision support tools such as Multi Criteria Decision Analysis (MCDA) to make decision making more transparent to both decision makers and to external stakeholders; 3) the opportunity costs of adopting a technology and whether the use of “cost-effectiveness thresholds” based on cost per QALY is appropriate; and, 4) whether decisions on the efficient use of a technology are adopted and reflected in *actual* use?

5. ***How should uncertainty as to the expected outcome be measured, conveyed to decision makers and used by them in decision making?***

All evidence and resulting decisions will have a degree of uncertainty, which presents challenges for both analysts and decision makers alike. One option to consider is the possibility of making conditional decisions – adopting the product while collecting additional evidence.

6. ***Are decision makers getting what they want in an efficient manner?***

There is growing interest in the question as to whether CEA is becoming “gold plated”, with more analysis being delivered and time taken than is necessary given the decision to be taken, the information available, and the opportunity cost, in terms of the necessary HTA resources available to look at other parts of the health care system.

Other issues are not addressed in this paper, particularly around governance processes (e.g., extent of stakeholder involvement, independence of assessment and decision-making processes) and the length of time take to undertake CEA reviews. For further discussion of these issues, see Drummond et al. (2009).

## 4. Valuing incremental health gain

### 4.1 Strengths and Weaknesses of the QALY

#### 4.1.1. Origins and Definition of the QALY

The QALY is a widely used measure of both quantity and quality of life. It was originally developed as a measure of health effectiveness for CEA so aiding decision makers charged with allocating scarce resources across competing health care programs (Fanshel and Bush, 1970; Torrance, Thomas and Sackett, 1972; Weinstein and Stason, 1977). The QALY is particularly useful because it enables comparisons across diseases, populations, and programs. As intimated above, a number of HTA bodies, such as NICE in the UK, use incremental QALYs as a key input to determine health care priorities. This original concept of the QALY has been called the “conventional” QALY (Lipscomb et al., 2009), recognizing that alternative conceptual models have been proposed, including an “equity-weighted” QALY.

Lipscomb et al. (2009) define what they term the conventional QALY as:

$QALY_{conv} = \sum_{t=1...T} \sum_{s=1...S} \rho_{st} V(H_{st})(1+r)^{t-1}$ , where:

- $\rho_{st}$  is the probability an individual will occupy health state  $H_s$  at time  $t$ ;
- $V(H_{st})$  is the value (or preference) measure assigned to the individual being in state  $H_s$  at time  $t$ ;
- $(1+r)^{t-1}$  is a discount factor designed to bring  $V(H_{st})$  to present value terms, with  $r$  being the selected discount rate reflecting time preference for health outcomes;
- $S$  is the number of discrete health states that may be occupied;
- $T$  is the time horizon relevant for decision making; and,
- $QALY_{conv}$  is subscripted to indicate this is (some variant of) the “conventional” QALY formulation.

The core concept of the conventional QALY ( $QALY_{conv}$ ) is grounded in decision science and expected utility theory<sup>6</sup>. The basic construct is that individuals move through health states over time and that each health state has a value attached to it. Health is defined as the value-weighted time—life-years weighted by their quality—accumulated over the relevant time horizon to yield QALYs. To permit aggregation of QALY changes, the value scale should have interval scale properties such that, for example, a gain from 0.2 to 0.4 is equally valuable as a gain from 0.6 to 0.8. States worse than death can be attributed a negative value and be subtracted from the number of QALYs. These conditions, along with an assumption of risk neutrality over life-years so that utility is additive over time, are sufficient to ensure that the QALY is a useful representation of health state preferences. These are very strong assumptions,

---

<sup>6</sup> For many extra-welfarists it does not need to be – see section 4.2. All that is necessary is that it is accepted as a credible measure of health for the purposes of social decision making.

but are necessary in order for QALYs to represent an individual's utility function for health over time. Violations of constant proportionality (in willingness to trade between quality and quantity of life) and hence of additive separability have raised issues around "end of life", which we discuss further later.

The following sections outline the four main stages or elements to estimate the change in QALYs associated with the use of a drug or other health technology.

#### 4.1.2 Valuing health states – which instrument to use?

First, the health state(s) ( $H_{st}$ ) of interest need to be classified and described to a respondent, who then values them relative to another health state, or an anchor point, such as death. The respondent then needs to value them by assigning preference values  $V(H_{st})$ , sometimes known as "utilities", because these can in principle be derived from a utility function. There are alternative approaches for eliciting the health state preference values, such as the standard gamble, the time trade-off, the visual analogue scale (VAS), and the person trade-off. Each approach has its own underlying rationale and, in some cases, set of axioms. For such comparisons to be helpful, different tools should yield much the same values for the same states. However, this is not often the case. Typically, the standard gamble yields higher values than the time trade-off, which in turn yields higher values than the VAS (Fryback et al., 2007). An alternative that is very widely used is a pre-scored health status instrument called a multi-attribute utility (MAU) measure. These include the EQ-5D, Health Utilities Index (HUI), SF-6D, and QWB. There are, however, important differences between MAU instruments. Part of the problem is designing an MAU with enough domains and levels to ensure there is adequate coverage along the full continuum of outcomes associated with each health domain. Fryback et al. (2007) demonstrate that these systems yield "similar but not identical trends" in health-related quality of life (HRQOL). Consequently, if the systems were brought to bear concurrently in a given economic evaluation, they would yield different QALY scores and thus possibly different conclusions about the cost-effectiveness of the intervention of interest.

#### 4.1.3. Valuing health states – whose values to use?

The second element is to decide whose values should be used as the source of valuation of the health status. Candidates include patients, decision makers, and the general public. The weights or value "tariffs" of all the MAU measures are based on surveys of the general public, although the EQ-5D also includes a VAS assessment of self-rated health for patients, which is often used in the context of a clinical study (Insinga et al., 2003).

The argument for using patient valuations rests on the belief that patients themselves are the best judges of the relative desirability of their own health states. However, the US Public Health Service Panel on Cost-Effectiveness in Health and Medicine recommended that a representative sample of the general population be used for health state valuations, as long as the judgments supporting these valuations are “informed, unbiased, and competent” (Gold et al., 1996). This recommendation is mirrored in NICE’s guidelines, based on the argument that in publicly funded health care systems, the aim of economic evaluation is not to make decisions at the individual patient level, but to guide policies that fulfill the interests of society as a whole. A further argument is individual patients could bias their responses because they have a vested interest in getting access to treatment. However, the practice of using valuations of the general population becomes problematic if the “informed, unbiased, and competent” requirement is not met – for example, because respondents have misunderstood what it is really like for patients to live with illness, both in terms of the impact on quality of life and the ability to adapt over time.

A comprehensive review of empirical studies in this area was conducted by de Witt et al. (2009). They analyzed 38 studies comparing patient and non-patient valuations and reported that of the 27 studies indicating at least some difference between the two groups, 22 reported higher (better) patient values; two reported lower patient values; and, three showed contradictory results. Hence, the majority of evidence indicates that, on average, patients tend to value their own health state more highly than do non-patients. This means that for interventions aimed at, say, restoring low HRQOL individuals to full health, the use of public valuations will generate *larger* health gains than if patient valuations were to be used. However, this also implies that interventions aimed at extending the lives of such individuals will be valued lower by the general population than by patients due to their perceptions about the level of HRQOL being maintained. Thus, the use of general population values will tend to favor interventions aimed at achieving perfect functioning, while disfavoring life-prolonging interventions.

If we accept the normative argument for using general population values, then it is important that respondents are well-informed about what the patient experience is actually like. For example, health state descriptions could be made more realistic by adding extra dimensions or by providing respondents with testimonials from patients with the illness in question. However, there will be a trade-off between providing sufficient detail and overburdening respondents. There is also some evidence that the provision of clinical information can induce an emotional and unconsidered response in general population respondents, which suggests that any hint that the health state relates to cancer (or some other high-profile disease) is likely to introduce bias. These factors indicate that the task of informing general population respondents about

the patient experience, and eliciting health state preferences more generally, is far from straightforward.

Concerns about the ability of the general public to value health states has led to more radical criticisms and proposals. For example, Kahneman (2007, 2009) argues that as people do not obey the axioms of expected utility, a paradigm shift is needed from decision utility, which underlies the QALY and is based on inference of choice, to “experienced utility” derived from behavioral economics (Dolan and Kahneman, 2008). “Experienced utility” attends to issues around patient adaptation to disease, which may better facilitate understanding of the differences between patient and public valuations of health states. It also suggests an overemphasis on the importance of health to general well-being brought about by asking people to focus on the link. Use of behavioural economics is a relatively new approach towards developing optimal methods of measuring patients’ experience.

#### 4.1.4. Valuing health states – profiling over time

The third element in the estimation of QALYs is the construction of the profile of health state values over time. In estimating the QALYs gained from a health care intervention, the analyst calculates the area under the curve; that is, the difference between the profiles obtained for the treatment of interest and its comparator. The simplest and most frequently used approach for estimating the profile is to multiply the time spent in each health state over time with its corresponding health state value. Nevertheless, some researchers argue that it is better to estimate the value of the whole profile  $V(H_{st})$ , because the value of a given health state  $H_s$  is not independent of the time spent in it, or the order in which it is experienced. One particular measure, Health Year Equivalents (HYE), has been put forward to address this problem (Mehrez and Gafni, 1989). However, it has been criticized (Culyer and Wagstaff, 1993) and has not been widely used because of the practical problems of obtaining the estimates.

#### 4.1.5. Valuing health states – concerns for fairness

The fourth element in estimating the total QALYs gained from a given health care intervention is the aggregation of QALYs across all the recipients. For example, should all QALYs be weighted equally, irrespective of who receives them?

It is widely assumed by health economists conducting CEA that the principal objective of health care is to maximize population health using available resources (Culyer, 1997). Given that the QALY has been developed to provide a generic measure of health effect, it follows that health care resource allocation should seek to maximize the number of QALYs generated (Dolan, 2001). This is commonly referred to as the QALY-maximization rule, and can be seen as a form



of ethical utilitarianism. Culyer (1992) refers to this position as “QALY egalitarianism” – all QALYs are of equal social value, regardless of to whom they accrue to and the context in which they are enjoyed. In other words, “a QALY is a QALY is a QALY” under all circumstances.

From a welfare economics perspective, health is not given a special status and the objective of health care is the maximization of social welfare. The use of an alternative extra-welfarist perspective (see Section 4.2) explicitly allows for distributional objectives beyond the pursuit of aggregate health gain (Williams and Cookson, 2006). Other ethical approaches discussed for distributing health care include achieving equal health or reducing health inequalities, suggesting that QALYs are of more social value when delivered to groups of patients disadvantaged in some way. A Rawlsian “veil of ignorance” ethical approach would put emphasis on maximising the health of the most ill individuals in society, putting an emphasis on disease severity. Lastly, we can note the “fair innings” approach (Williams, 1997), which argues that everyone is entitled to some common quantity of life time QALYs. Potential QALYs to be gained by those who would otherwise not enjoy a “fair innings” are worth more than QALYs gained by older people who have already had their “fair innings.”

Three particularly salient ethical problems can be identified. First, the standard QALY approach gives no regard for the pre-treatment utility level of the individuals concerned. Both ethical theory and public opinion in a number of industrialized countries suggest that in setting priorities, society emphasizes how badly off the individuals would be if intervention did not take place. That is to say, the worse off an individual would be without an intervention, the more highly society tends to value that intervention. This aspect of societal valuation is often referred to as an independent concern for severity. Empirical studies of severity-related preferences suggest that people are, on the whole, willing to sacrifice aggregate health in order to give priority to the severely ill. In quantitative population preference studies (see Ubel et al., 1999; Richardson, 2007; Dolan et al., 2005; Nord, 1999; Shah, 2009), QALY gains to severely ill groups have been weighted two to 10 times more highly than gains to less severely ill groups. The person trade-off (PTO) offers an explicit community weighting of QALYs according to disease severity (Nord, 1995)<sup>7</sup>.

Second, the conventional QALY model implies that the value of an intervention is proportional to the beneficiary’s capacity to benefit. The model therefore favors those with more treatable conditions and those with greater potential for health, in terms of functioning and/or longevity. This is somewhat at odds with both ethical theory and public opinion, which suggest that it should not be held against people that they happen to have conditions for which there are no complete cures or that their remaining lifetime is somewhat limited. In a UK study, Dolan and Cookson (1998) found a reluctance to discriminate in situations where groups differed in terms

---

<sup>7</sup> In a QALY TTO valuation individuals trade their own willingness to live in a state of ill health for a smaller number of years of full health. In a PTO valuation individuals are also required to trade between two states but in doing so (i) take into account not only the outcome of the intervention (the change in health status) but also the starting point (severity) and (ii) make a *social* judgment about how many people they would be willing to stop (trade) treating in one case in order to treat a given number of patients in the other case.

of potential for gaining life years (e.g., 10 vs. 20 years). People seem to believe that there should be priority accorded to those in urgent need of medical attention, and that while capacity to benefit does matter, it is a secondary consideration.

The third problem is a special case of the second issue. Valuing health gains in terms of QALYs means that life-years gained in full health—through, for instance, prevention of fatal accidents in people in normal health—are counted as more valuable than life-years gained by those who are chronically ill or disabled (e.g., averting fatal episodes in people with asthma, heart disease, or mental illness). This also runs counter to results obtained in studies of public preferences for priority-setting.

All of these issues arose in a debate in the UK over therapies given at the “end of life”. NICE was asked to appraise four drugs for advanced or metastatic renal cell carcinoma. Following an assessment of clinical and cost-effectiveness, NICE concluded that none of the four drugs were cost-effective. As a result of the extensive debate following the decision, NICE issued supplementary guidance to the Institute’s Appraisal Committee for therapies given at end of life (NICE, 2009). If a therapy, for a condition affecting a small patient population with less than 24 months to live, adds three months or more life extension, then the Committee should: 1) assume that the QALYs gained are at full quality of life in the added months and 2) consider that the QALYs gained be given a much higher weight relative to NICE’s current threshold<sup>8</sup>.

The methodological and practical issues in the derivation and use of QALYs have been discussed in a supplement to the journal *Value in Health* (Drummond, 2009), which is a useful source for further details.

#### 4.2. Extra-welfarism versus welfarist use of the QALY

Garber and Phelps (1997) developed a traditional welfare economics model of optimal purchase of medical care by an individual paying out-of-pocket and showed that, under certain assumptions, the individual’s marginal or threshold willingness-to-pay for medical care can be expressed as an incremental cost-per-QALY gained. In other words, individuals would consume health care until the point at which incremental benefits equaled opportunity cost.

Culyer (2008, 2009), however, has three main criticisms of such a “welfarist” approach to resource allocation in health. The first issue regards the focus on individual utility derived from preferences over goods and services. Extra-welfarism takes a more comprehensive view of well-being. In the context of delivering health care, it is the ability of an intervention to improve health status (including functioning and psychic state) and/or the length of life that matters, not the utility value. This aspect of extra-welfarism draws on Sen’s theory of capabilities (Sen

---

<sup>8</sup> The argument for an end of life uplift to give more value to the life extension is the violation of the constant proportional trade-off assumption that underlies the conventional QALY (as referred to in section 4.1.1 above). The willingness to accept a higher threshold reflects society’s desire to place an emphasis on the severity of the condition being treated. There is no obvious rationale for the restriction to “small” (undefined) patient groups.

1987a, 1987b, 2002; Cookson 2005; Mason and Towse, 2007) and arguably reflects a tradition developing on views on “merit goods” within public finance (Sugden, 2005, 2008). Secondly, he takes issue with the use in the standard welfarist approach of any improvement in wellbeing requiring a Paretian (or Kaldor–Hicks) improvement in order to avoid interpersonal comparisons of utility. Interpersonal comparisons are essential for decision making about resource allocation in health care. Thirdly, the application of willingness-to-pay techniques has limitations. It does not help in valuing social preferences for health care resource use. Its use to establish preferences over individual health states raises income inequality and therefore potential equity issues.

Although CEA at the level of the individual, and arguably at the third party payer level, can be grounded in welfare economics, the use of QALYs is usually implicitly or explicitly used within an “extra-welfarist” approach. This approach is to support the decision maker acting on behalf of the third party payer to allocate resources within a set of objectives that may have come via a political process and are not based on, or limited to, utilities derived from individual’s preferences over goods and services.

It is, however, sometimes difficult in the welfarist versus extra-welfarist debate to clarify whether the debate is about the answer to an empirical question as to which approaches are most helpful to decision makers, or whether the debate reflects an ideological difference around the role of markets in the allocation of health care resources. This includes the important question as to how equity issues are taken into account, such that people on low incomes or with great needs get a “fair” share from society’s viewpoint (Brouwer et al., 2008).

#### 4.3 Alternatives to the QALY

Some HTA bodies – most notably IQWiG - have responded to criticisms of the QALY by actively pursuing disease-specific alternatives<sup>9</sup>. This raises the question as to whether it is possible and useful to depart from a generic metric and some sort of common threshold measure of opportunity cost. In our view there are four options:

1. Continue to use the QALY, while seeking to improve its ability to measure health status and methods to value that health status.
2. Move to WTP and CBA based approaches valuing health gain in monetary terms, which we discuss below.

---

<sup>9</sup>Disease specific measures are usually more sensitive, but, a payer comparing two drugs in two different disease areas using two different disease specific is “comparing apples and oranges.” A common numeraire is needed to inform resource allocation or pricing decisions. The approach of IQWiG, based on an “efficiency frontier”, has been much criticized (Caro et al., 2010; Jönsson, 2008; Brouwer and Rutten, 2010; Dintisios and Gerber, 2010; Schulpher and Claxton, 2010).

3. Use some form of hybrid. For example, IQWiG is in principle seeking to use disease-specific instruments with interval scale properties<sup>10</sup> in combination with the use of WTP studies to understand the importance attached to particular disease areas by German citizens<sup>11</sup>.
4. Revert to a pre-HTA world in which effect and cost are addressed by different parts of the health care system, which in the case of pharmaceuticals has involved the use of a variety of price and cost control mechanisms of the types discussed in papers 8 and 9.

Our view is that Option 1 makes sense, but analysts should explore the potential for Options 2 and 3 to offer alternatives, or, more realistically, routes to improve the QALY-based approach.

We also note that the World Health Organisation uses Disability Adjusted Life Years (DALYs) in its approach to CEA (Generalised Cost-Effectiveness Analysis), rather than QALYs (Tan-Torres Edejer et al., 2003). DALYs like QALYs combine length of life with a measurement of health states. They use a standardized life expectancy, PTOs. to value health states, and apply different weights to years lived at different ages. For an exposition see Murray (1994) and for a critique see Fox-Rushby (2002).

#### 4.4 The use of WTP methods to value health care interventions

The last 15 years has seen a surge in the number of methodological applications in the area of contingent valuation and the measurement and valuation of monetary benefits, with the increased use of both WTP and stated preference discrete choice experiments (SPDCEs) for estimating values in areas such as environmental and transport economics (Donaldson et al., 2006). These methods use surveys to elicit people's preferences for public goods by finding out what they would be willing to pay (accept) for specified improvements or down-gradings in them. This can then be compared with the cost of the good, within a CBA.

However, in order to meet these aims, as Mitchell and Carson (1989) note, "the survey must simultaneously meet the methodological imperatives of survey research and the requirements of economic theory". To this end, any scenarios used in surveys must be understandable and meaningful to the respondents and free of incentives that might bias the results. The majority of the criticisms of survey research are associated with the possible biases, which are a threat to the validity of WTP results.

Lack of consensus as to how to conduct health care contingent valuation studies is reflected in the wide variation amongst studies, in terms of the types of questions being posed and the elicitation formats being used explored (Diener et al., 1998). Drummond et al. (2005) note, "most of the published health care contingent valuation studies are experimental in nature,

---

<sup>10</sup> Most disease specific outcome measures, e.g. reduction in cholesterol, do not have interval properties. A 20% reduction in cholesterol is not four times better than a 5% reduction.

<sup>11</sup> The same result could be achieved (probably a lot more easily) by measuring incremental outcomes in QALYs and using disease specific cost-per-QALY thresholds which could be informed by WTP.

attempting to explore measurement feasibility issues rather than being full programme evaluations using CBA". To address some of the variability in methods, the NOAA panel (Arrow et al., 1993) carried out an assessment of the various approaches to eliciting WTP. Their recommendations included promotion of the closed-ended format, face-to-face interviews instead of surveys, pilot surveying and pre-testing and provision of accurate information on the good being valued. However, there remains no consensus on the best approach in health care. There is still a requirement for methodological work in the area of WTP surveys, and in the use of contingent valuation methods more generally. As noted earlier, there are preference elicitation issues in QALY-based CEA methods as well.

SPDCEs are not necessarily an alternative to QALYs. They can be used to derive WTP values for QALYs. Where third party payers compete for premiums from enrollees SPDCEs can in theory help to identify how much additional premium (ex ante) enrollees are prepared to pay for access to particular technologies and the benefits they bring as measured in QALYs. SPDCEs can, in theory, be used to construct QALYs, for example by translating (trading) disease specific effects into effects recognised by a MAU QALY measure. As we note in a later section, they can be used to derive the social value of a QALY. They can also help identify and incorporate into a CEA non-health attributes that are valued by patients (such as greater convenience, over and above any impact on compliance and productivity), and are being used to help regulators understand patients willingness to trade-off the risks and benefits of drug treatment (Cross and Garrison 2008).

## 5. Which costs and benefits to include alongside health gain?

As discussed, CEA usually includes the QALY benefits to the patient arising from the intervention and all of the costs to the health care system of the treatment over a suitable time horizon discounted as appropriate. A societal perspective, as more conventionally used in CBA outside of health care, includes *all* costs and consequences related to the initial interventions (Jönsson, 2009). This includes unrelated medical costs<sup>12</sup>, costs incurred outside the health care sector, and benefits accruing to all stakeholders in society including those for the patient not captured in the QALY. The question arises as to how many, if any, of these other factors or considerations should be included (Drummond et al., 2008)? Several countries, including Australia, Norway, Sweden and the Netherlands, already require that economic evaluations are conducted using a societal perspective. But what exactly is included varies.

It is helpful to separate out the main elements of benefit and cost. Health benefits for patients and direct treatment costs falling on the health care system should always be included in CEA. There is growing recognition that unrelated medical costs that might fall on the health care system at some future point should (at least in principle) be included in a CEA (Weinstein and Manning, 1997). If a treatment extends life for example, then patients will need more health care in the future for unrelated illness. These benefits/costs are included in the Swedish and Dutch guidelines. There is greater controversy as to whether or not to include wider consumption and earnings effects, such as the productivity or earning benefits of getting people back to work; impact on carers, other public programs, and on consumption patterns; and, any out-of-pocket costs incurred by patients linked to obtaining treatment.

Meltzer (1997) argued that CEA has to take into account all costs and consequences related to the initial intervention. Here the starting point is (as with Garber and Phelps, 1997 discussed above) an individual working for earnings and buying health care along with other goods. The individual wants to maximize expected lifetime utility. To have a long and healthy life, they have to spend money on health care and save during their working lives to enable them to achieve this. For the individual, the lifetime constraint is therefore their total budget. There is no separate constrained health care budget. However, for the point of view of a government operating a constrained health care budget as a single (third party) payer, seeking to maximize overall utility (i.e., the value of health gain relative to additional expenditures on consumption and medical care net of any earnings) could lead to choices that are not acceptable to society. For old people, on average, consumption exceeds earnings, so interventions that prolong life into old age are of less value to society than those that get people of working age back into employment<sup>13</sup>. This problem can be avoided if time gained (be it additional years of life, or time not spent with a disease, or providing informal care for an ill person) is regarded as of equal value, however the time might be used. However the “fair innings” argument of Williams (1997) suggests that “a year is not a year”, i.e. age is a factor influencing the value of a QALY.

---

<sup>12</sup> “Unrelated medical costs” are those that will arise because the patient lives longer and therefore can be expected to require medical attention for other diseases during this added period.

<sup>13</sup> Of course a weighting in favour of working age life years is explicitly built into the DALY as we discuss earlier.

Claxton et al. (2010) have argued, from an extra-welfarist perspective, that the holder of a constrained health care budget should not take broader costs and benefits into account in the absence of off-setting resource transfers into the health care budget to compensate for purchasing extra non-health benefits or reduced non-health care costs<sup>14</sup>. The issue comes back to the purpose of health spending, the importance of health as a social good, and how the health care budget is set. For instance, in a public sector system will the fact the health spend is delivering non-health goods lead to a higher health care budget? In a competitive private insurance system, will enrollees or those paying premium on their behalf be willing to pay higher premiums to cover interventions that deliver valuable non-health benefits? A US employer may be interested in interventions that make its workforce less likely to be absent, but less interested in reducing potential burdens on family members not on its payroll.

---

<sup>14</sup>They also argue for an approach that would downrate the broader costs and benefits to reflect the constraint on the single payers health budget (as reflected in the ratio of the single payer's opportunity cost value of a QALY to citizens willingness to pay for a QALY).

## 6. What types of evidence to use?

### 6.1 Economic evaluation in drug development

Economic evaluation plays a role in all phases of drug development. Phase 1 trials usually consist of administration of single, conservative doses to a small number of healthy volunteers. The effects of increasing the size and number of daily dosages are evaluated in order to determine the point at which the likely therapeutic dosage has been exceeded. Cost-of-illness studies are carried out during this stage in order to aid decisions on priority setting, whether to further develop the drug, and to gather background data for future economic evaluations.

In Phase 2 clinical trials, a drug is administered to limited number of patients with the target disease, in order to explore its potential therapeutic benefit. At this stage, efficacy is still uncertain. Again, cost of illness studies help to determine the potential market for the drug and lay the groundwork for formal economic evaluations. In addition, some of the instruments for gathering economic and quality of life data may be piloted during this phase.

Related applications of Phase 1 and 2 economic evaluations also include guidance of decisions to focus further development in specific indications or patient groups; strategic positioning in relation to competitors; and, determination of optimal sample size and data elements to collect prospectively during the conduct of Phase 3 studies. Phase 3 studies involve large, often multi-center, randomized controlled clinical trials designed to determine efficacy. Such trials typically involve careful patient selection with clearly specified inclusion and exclusion criteria. At this stage of drug development, formal pharmaco-economic evaluation is indicated in order to determine the cost-effectiveness of the new drug. Such evaluations are often conducted alongside the randomized controlled clinical trials and commonly used in reimbursement applications.

Economic data during the various phases of drug development are either gathered prospectively as part of the trial protocol or retrospectively following a trial or through modeling using data from various sources. However, there remain several methodological issues that relate to:

*Atypical nature of trial settings:* Clinical trials are usually performed in specialist centers by highly committed investigators enrolling selected patients who have a higher likelihood of complying with therapy than patients in usual care. Also, patients in the trials are usually selected based on strict inclusion and exclusion criteria that would then limit generalizability of findings (e.g., exclusion of pregnant women, children, and very elderly people).

*Protocol-driven costs and benefits:* Clinical trial protocols often include close monitoring of enrolled patients through pre-scheduled extra visits and tests. Protocol-defined testing for efficacy or safety purposes will increase the costs of diagnostic testing in the trial, whereas many of these tests would not be performed in usual care. Protocol-driven tests or procedures



may also induce detection of extra cases or adverse events that would otherwise have gone undetected. In addition, detected outcomes may be treated more aggressively than they would be in usual care.

*Inappropriate clinical alternatives:* Trials usually compare the new therapy with a placebo or baseline therapy which may not represent the most appropriate alternative in current clinical practice.

*Inadequate length of follow-up:* Almost all trials use intermediate or surrogate endpoints such as disease progression or biomedical markers. Ideally, economic analysis should be based on final outcomes related disability or mortality.

*Inadequate sample size for economic analysis:* Sample size calculations in trials are based on the clinical endpoints which may not be sufficient for highly skewed economic parameters (e.g., length of hospital stay, total cost).

*Inappropriate range of end-points for both costs and consequences:* Ideally, economic analyses require data on resource use and outcomes of direct relevance to patients (e.g., quality of life).

Several methods are being investigated to overcome these problems. For example, in situations where no head-to-head clinical trials exist comparing the alternatives of interest, it may be possible to use indirect and mixed treatment comparison approaches for evidence synthesis. These approaches have great potential for estimating the comparative effectiveness of multiple treatments using an evidence base of trials that individually do not compare all treatment options. Connected networks of evidence can be synthesized simultaneously to provide estimates of the comparative effectiveness of all included treatments and a ranking of their effectiveness with associated probability statements (Sutton et al., 2008).

Large and simple outcome trials (prospective randomized but pragmatic trials) deal with the problems of artificial settings and protocol-driven costs, but they are very expensive to conduct and can only be done in late Phase 3 or 4, once the efficacy of the drug has been established. However, during Phase 4 of drug development (the post-marketing stage), data can be gathered in support of the use of the drug in routine clinical practice. Post-marketing economic evaluations are critically important, as they allow the study of the costs and consequences of drug therapy outside the strict constraints of Phase 3 controlled clinical trials. In other words, economic evaluations in Phase 4 studies typically address real world costs, effectiveness and cost-effectiveness issues. This may in part be done through an RCT that compares the drug with current therapy, and may also be more pragmatic in design. However, clinical trial-based economic evaluations are being increasingly complemented by utilization and effectiveness data from observational studies which often use registries.

National methods guidelines vary in the role they see for estimates of clinical efficacy and effectiveness. Some guidelines, such as those proposed by IQWiG, are happy to see these estimates used as measures of benefit in economic evaluations. On the other hand, the

guidelines proposed by NICE require the health gain to be estimated in QALYs. In this case, the systematic review of RCTs becomes just one input, albeit an essential one, to the evidence considered. For example, observational studies may be required to help convert an intermediate outcome (e.g., change in HbA1c or LDL cholesterol) to a final outcome (e.g., change in coronary heart disease events or survival). In addition, estimation of the QALYs gained will require information on the effect the therapy has on the patient's quality of life. Therefore, decision-analytic modeling is usually required to produce an estimate of economic benefit in cost-effectiveness studies. The decision model provides a framework within which data from a number of sources, including clinical trials, observational studies and other routine sources can be synthesized.

It is worth noting that the two objectives of collecting clinical data, for drug licensing and for reimbursement, are sometimes at odds with one another. A study designed for one purpose may not be suitable for the other. Even the use of decision-analytic modeling may not obviate the problem. A case in point is in the development of anti-cancer drugs. Drug licensing agencies are happy to accept data on progression-free survival or time to progression as evidence of efficacy. If trials are stopped at the point where progression-free survival has been demonstrated, allowing patients to cross-over from the active treatment arm, this will reduce the chances of demonstrating overall survival, even if the patients are followed for a further period of time. Typically, the reimbursement agency wants to see evidence of overall survival (Drummond et al., 2009a).

Tensions between licensing and reimbursement data requirements have led to a growth in "early engagement" activities between companies and HTA/P&R agencies. Experiments of three way engagement between the company, the licensing agency and the HTA or reimbursement agency, are taking place in the UK, Sweden and Australia.

## 6.2 Issues in the transferability of economic evaluations

As a growing number of jurisdictions request economic data in support of their decision-making procedures for the pricing and/or reimbursement of health technologies, demands on study sponsors and researchers increase, especially as the various national guidelines may insist on the presentation of local data, or the use of specific methods (Barbieri et al., 2010).

There are many reasons why the cost-effectiveness of health technologies might vary from place to place, including the incidence and severity of the disease in question, the availability of health care resources, clinical practice patterns and relative prices (Sculpher et al., 2004). The extent of variation in estimates has been shown in a review of economic evaluations of medicines undertaken in Western Europe (Barbieri et al., 2005). They found that in 17 out of 27 cases the variation in the estimates of the incremental cost-effectiveness ratios could be considered to be substantial (i.e., a two-fold difference, which had the possibility to change the decision to reimburse the drug). Therefore, it is reasonable for national guidelines to request that analyses be relevant to the local context.

However, the requirement that economic evaluations should use local data, or that particular methods should be used, means that analyses increasingly need to be customized for each setting. Transferability is a key issue (Kneis, 2009). Drummond et al., (2009b) recommend good research practices for dealing with aspects of transferability, including analytic strategies and guidance for considering the appropriateness of evidence from other countries.

They define economic evaluations as *generalizable* if they applied, without adjustment, to other settings and *transferable* if they *could be adapted* to apply to other settings. The generic term “jurisdiction” was used to mean any setting where there is a need for local estimates of cost-effectiveness. Often this would be a country, but could also be a region within a country, or a particular payer, such as a health plan.

They confirm that there are several important methodological and practical issues surrounding the transferability of economic data. There is some evidence of the variability in the cost-effectiveness of health technologies between locations. Many international guidelines for economic evaluation make references to problems concerning economic data transferability and include requirements for jurisdiction-specific data or methods. Drummond et al. (2009a) recommend, however, that those developing guidelines for economic evaluation should fully justify the need for local data or methods, since this increases the burden on those undertaking studies in multiple jurisdictions and ultimately pushes up the costs of developing new technologies.

### 6.3 Towards an International Standard for Economic Evaluation

It is less clear why methods vary. Drummond and Rutten (2008) made a proposal for an international reference case<sup>15</sup> for economic evaluation. The question is which elements of methods should be standardized and which could legitimately vary by jurisdiction. For example, the principle that costs occurring in future years should be discounted to present values should be an international standard, but the discount rate could legitimately vary by jurisdiction. Or, a societal perspective could be required with each cost and benefit separately identified. Jurisdictions would choose which subsets of these costs and benefits to include depending on the context and therefore perspective of their decision makers.

If the methodological requirements of different jurisdictions were more similar, this would reduce the burden on those making submissions to reimbursement bodies in multiple countries. Attempts to standardize methods will probably be the most fruitful forms of international collaboration. The main attempt to date at increasing international collaboration

---

<sup>15</sup> A “reference case” was put forward by Gold et al. (1996). It combines (i) a standard set of methods for conducting economic evaluation, including the specification of acceptable study design and data sources, with (ii) a standard way of reporting the study. Other countries have implemented this approach, for example, NICE introduced its “reference case” in 2004 methods guidelines, updated in 2008 (NICE, 2008).

is the EUnetHTA project ([www.eunetha.net](http://www.eunetha.net)) promoted and funded by the EU designed to produce a “core” HTA template for an assessment that could potentially be used by decision makers in several jurisdictions.

Progress will depend on the extent to which there is agreement on common methods and requirements and on the extent to which evidence in one jurisdiction is relevant to another. In other words, real benefits from international collaboration may come not from methods convergence but if some data is transferable with little or no adaption required. Trueman et al. (2009) examine the feasibility of harmonizing HTAs across jurisdictions, based on a case study of drug eluting stents, which was the first application of the EUnetHTA core model. They argue that, although there is a core common data set considered by most of the HTA agencies, differences in the approach to HTA, heterogeneity of studies, and the limited relevance of findings to local practice meant that the data set that came from the core model would only have had a limited influence on the resulting national recommendations.

The potential for common HTA assessments (using agreed methods and data generalizability) is greater in the systematic reviews of the clinical efficacy data than in the economic evaluation component. EUnetHTA is exploring the potential for single relative effectiveness assessments to be shared by Member State HTA bodies. However, methods for conducting reviews and willingness to use indirect comparisons differ (Jönsson, 2011).

There have been several bilateral and trilateral arrangements between HTA bodies across countries, for example between HAS, IQWiG and NICE. Whether as a result of this collaboration or not, there has been some convergence of methodological requirements over time. For example, NICE has recognized that equal weighting of QALYs may not be appropriate in all situations. IQWiG guidance recognizes the need to combine the various outcomes of health care programs, in a measure akin to a QALY, and also allows the modeling of both costs and benefits over the patient’s lifetime, an approach similar to that used by NICE. As more and more jurisdictions begin to require HTAs as part of their process for deciding on the reimbursement or coverage for new health technologies, the case for international collaboration increases.

The final step in international collaboration could be to make a common decision, as is currently the case for drug licensing within the EU. Some have argued for a “Euro-NICE” making decisions on behalf of all EU Member States. There are scientific, practical, and political issues here. As noted in 6.2 above, clinical practice may vary, with implications for the choice of relevant comparator and resource use. Relative effectiveness will depend on patient mix, baseline risk, and the comparator. Cost-effectiveness is likely to vary between countries. Even where it does not, decision makers in one country, faced with the same assessment of evidence, may still come to a different decision than those in another jurisdiction about use of the same technology. This is because countries have different levels of resource to devote to

health care and different priorities. Since international prices of new health technologies do not vary greatly<sup>16</sup> it is likely to remain the case that some countries can afford them and others not.

## 7. Making decisions

### 7.1 Weighting different factors in a deliberative process

Most HTA bodies have a committee to assess the evidence and make a decision. The mechanism by which the members of a committee combine the various forms of evidence with local context and judgements about interpretation and uncertainty to reach a decision is a deliberative process (Culyer, 2009). Advocates for the positive value (as opposed to necessity) of deliberative processes argue that (a) they increase the likelihood of achieving satisfactory decisions and (b) the participatory and consensus-building nature of such deliberations is likely to have a positive impact on the perceptions of the democratic process, so long as publicity and transparency (Daniels and Sabin, 2002; 2008) are maintained. Culyer and Lomas (2006) suggest that deliberative processes may be particularly valuable in circumstances where there is either uncertainty about technical information, or where issues relating to fairness and social values need to be taken into account.

Culyer (2009) has described the methods used by NICE for evaluating health technologies as a good model of a deliberative process. NICE holds extensive consultative exercises through the appraisals process and synthesizes opinions and information from a large number of stakeholders including manufacturers, health professionals, patient groups and lay representatives. Appraisal Committee membership is set to ensure broad representation. NICE's processes have attracted praise from the WHO (Hill et al., 2003) for "use of best available evidence in decision making, transparency, consultation, inclusion of all key stakeholders, and responsiveness to change."

Whilst NICE appears to be more transparent than other HTA bodies, some researchers are critical of its failure to formally codify the impact of decision criteria other than cost-effectiveness, claiming that its statements on these matters have been vague and uninformative (Schlander, 2008). The importance of social value judgements and other factors beyond cost-effectiveness is regularly emphasized, and examples of interventions with high questionable cost-effectiveness being recommended on the basis of such factors are given (Rawlins et al., 2010). These are, however, unusual cases. It is difficult in most cases to understand the extent to which they have contributed to the final recommendation decisions and it is not possible from a review of decisions to find any factors other than the threshold that explain NICE decisions<sup>17</sup> (Devlin et al., 2011). Thus, while in principle NICE's decision

---

<sup>16</sup> Differential pricing would be more efficient.

<sup>17</sup> There is, however, evidence that a higher threshold is used for cancer treatments. This pre-dates the "end of life" guidance.

making fits the description of a sound deliberative process, the lack of explicit reporting of this process means that clarity is not always achieved.

This raises the question as to whether decision support tools can improve the transparency and effectiveness of a deliberative process.

## 7.2 Using decision support tools such as MCDA

Multi-criteria decision analysis (MCDA) methods have been advocated for use in health care priority-setting (Devlin and Sussex, 2011; Dowie, 2008). MCDA is a methodology for appraising options on multiple (often conflicting) criteria with the goal of providing a combined appraisal that includes an overall ordering of those options. It provides a framework for explicitly trading off various objectives against each other. It is particularly useful when these objectives do not share a common unit of valuation – for example, health care programmes typically involve a mixture of health, monetary, distributional and political objectives. MCDA has been proposed to support the benefit-risk assessment process conducted for making drug regulatory decisions (i.e., marketing authorizations or withdrawals for drugs) (Cross and Garrison, 2008) and the EMA is researching the use of this approach. MCDA is used routinely in other disciplines such as environmental and agricultural sciences (Baltussen and Niessen, 2006). For example, in the analysis of candidate UK nuclear waste disposal sites, MCDA was found to be superior to the Government's existing process, which had failed to include all key stakeholders objectives and had not made all value judgements explicit (DLTR, 2000).

In the context of a CEA decision, the decision-making committee could identify a list of relevant criteria for decision making (e.g., cost per QALY ratio and patients' disease severity measured by proximity to death). The relative weights for the various criteria could be generated from the deliberations of the committee or initially be introduced from elsewhere and the committee asked to review. The committee would generate scores (e.g., a weighting to achieve a "tackling inequality" objective) relative to the product in question based on the available evidence and judgement. This information could be incorporated into an MCDA model to obtain the total weighted scores. The committee could review and change the weightings. Sensitivity analyses would help to explore the effects on the overall results of differences in the scores and weights (Phillips and Costa, 2007). The committee would control the decision by altering the weightings. The analysis could then be reported in a final decision statement which identified those factors that had been important.

Use of MCDA would be attractive if it led to processes becoming more transparent and systematic. The burden on decision makers of using this approach would need to be proportional.

## 7.3 The opportunity cost of adopting a technology

Decision makers need to know what they are giving up if they adopt a drug. In the case of a health care system with a "hard" fixed global budget, such as the UK, the opportunity cost is

usually displacing another health care-related activity. In the case of a “soft” public budget system or a private sector system, adoption may lead to increases in taxes or premiums and so reductions in private consumption elsewhere.

We consider four issues in the context of a global “hard” fixed budget, using the UK NICE as an example.<sup>18</sup>

Firstly, should the threshold be explicit? In its early days, NICE denied that it was applying a specific threshold. However, as the information on the decisions made by NICE accumulated, it was possible to estimate a revealed threshold (Devlin and Parkin, 2004). NICE then stated that it applied a threshold range; interventions with an incremental cost per QALY ratio below £20,000 have a high probability of funding, those with a ratio exceeding £30,000 have a low probability of funding (Rawlins and Culyer, 2004) although the upper bound of £30,000 can be exceeded, for example on grounds of equity. The NICE Methods guide (NICE, 2008) refers to this threshold range.

The main arguments for an explicit threshold are that it is transparent and may encourage more consistency in decision making. Also, as happened with NICE, a threshold would be inferred even if it were not explicitly stated. The main arguments against an explicit threshold are that it may tell us little about the real opportunity cost of adopting a new technology (Birch and Gafni, 2006) and that it provides guidance to technology manufacturers on the maximum amount they can charge. This latter point, however, is arguably positive. Manufacturers know what payers want to reward and invest in R&D accordingly. As an indirect price control it is as likely to constrain prices as to encourage companies to raise them. For an early discussion of these issues see Drummond et al. (1997). We return to this issue below.

Secondly, how should the threshold be set? In order to apply an explicit threshold, the decision maker needs to know what the right level would be. NICE has never claimed to know the answer to this question. Some of those close to NICE have described the Institute’s decision-making process as a search for a threshold (Culyer et al., 2007; McCabe et al., 2008). Research in the UK is beginning to tackle the issue of the threshold level. In one study, Appleby et al. (2009) using case studies found it difficult to identify what in practice was displaced at local level. Another study, Martin et al (2007) attempted to estimate the threshold level implied by a cross sectional analysis of the current pattern of expenditure by disease area by geographic area within the NHS. Their conclusion is that NICE’s threshold range may be a little too high.

Thirdly, there are issues as to whether such a cost-effectiveness threshold encourages the “right” amount of innovation. Jena and Philipson (2008) have argued that the amount of the social benefit going to innovators will be too low. Others (Claxton et al., 2009) have argued that allowing innovators to price up to the threshold means they can appropriate all of the benefit.

---

<sup>18</sup> We do not consider budget impact per se (i.e. the incremental impact on the budget relative to its overall size). For a discussion of the role of budget impact analysis see Neumann (2007). For a discussion of the relationship between budget impact and uncertainty see Sendi and Briggs (2001).

Danzon et al. (2011) argue that the patent system is intended to provide temporary monopoly rights that can enable innovators to exercise some market power. The most dynamically efficient outcome is for as much as possible of the social surplus to accrue to the innovator in that period. Payer use of cost-effectiveness thresholds will be (second best) efficient if thresholds reflect societal willingness to pay. Jena and Philipson (2009) express concern, however, that if threshold constrained prices do not reflect relative marginal costs, then there may be a misallocation of resources. Lakdawalla and Sood (2005, 2009) argue that a two-part tariff is more efficient than prices set using buyer thresholds, because copayment can then be set at marginal cost to get optimal use.

Fourthly, should different thresholds apply, depending on the nature of the treatments or patient populations being studied? In part, this links back to the discussion above; if the QALY does not fully capture all the relevant elements of social value, it may not make sense to apply a single threshold. This issue has been raised in the context of drugs for rare diseases (i.e., orphan drugs). Even if these treatments do not appear very cost-effective (i.e., have a very high cost-effectiveness ratio), society may still prefer to make them available, because many of the diseases treated with orphan drugs are life-threatening or because it would be unfair for someone not to be offered treatment just because their disease is rare (Drummond et al., 2007). On the other hand, it has been pointed out that valuing rarity *per se* may not make sense, in that resources could be denied to patients suffering from more common diseases, some of which could be equally serious (McCabe et al., 2005)<sup>19</sup>.

Further research is needed on whether QALYs should be weighted other than equally and whether society values attributes of health care that cannot easily be incorporated in the QALY framework. It is possible that factors such as the severity of the patient's condition, or the availability of alternative treatments, could be relevant, but there is little evidence as to how these feature in decision making by HTA bodies. An exception is NICE's advice when appraising end-of-life medicines with cost per QALY ratios exceeding £30,000 (NICE, 2009).

In "soft" public budget systems and private sector systems, where more health care spending will displace private consumption then societal "willingness to pay" for health gain is the more appropriate measure of opportunity cost. Efforts in the UK to assess this led to a major research effort launched to determine the feasibility of estimating a monetary value of a QALY from a societal perspective (The SVQ Research Team, 2008; Mason et al., 2008; Baker et al., 2010). Results from this research estimate societal WTP for a QALY in the UK ranging from £70,000 per QALY for immediate life-saving, through £35,000 per QALY for life-extending over time, to £10,000 per QALY for quality-of-life-enhancing effects. On this basis, NICE's threshold range would roughly double in some cases, but would fall for treatments that did not affect life expectancy.

---

<sup>19</sup> Orphan drug legislation requires diseases to be severe as well as having small patient populations. The real issue is therefore the nature of the social weighting that should be applied.



## 7.4 Implementation of cost-effective treatments

Successful dissemination and implementation of treatments determined to be cost-effective is a challenge for health care systems. Less attention has been paid to the implementation of decisions or guidance. This creates a significant challenge for HTA bodies, as well as for local decision makers and practitioners. In the UK, for example, an analysis of public comments on NICE suggests that there is a significant concern regarding the patchy and slow implementation of adoption recommendations, with many stakeholders deeming this a key issue in terms of the Institute's effectiveness, efficiency, and public credibility (Sorenson et al., 2008).

There are a range of issues that influence whether recommended treatments are indeed used in health care systems. Such factors include insufficient or misaligned policy aims (i.e., differences in objectives between the HTA process and the needs of decision makers – the more decentralized a given health system, the more this may be an issue); a lack of a holistic approach to implementation, where not all relevant stakeholders are informed of decisions or there is poor dissemination of guidance; limited use of formal mechanisms to enforce implementation; and, rapidly changing political situations (Hutton et al., 2006; Audit Commission, 2005; OECD, 2005; Neumann, 2004; Sheldon et al., 2004). In addition, local authorities often deal with different resource capacities, patient populations, health needs and available budgets, which can impact their ability to implement national decisions or guidance to make treatments accessible to their populations. A study of NICE guidance implementation by the Audit Commission (2005) found that poor financial planning by local health authorities, in terms of adequately estimating the costs and resource requirements of implementation, was one of the leading factors contributing to poor implementation. While negative guidance will always be implemented, positive guidance often may not (Drummond et al., 2009a), due to the resource consequences and difficulty making disinvestment decisions elsewhere. Buxton (2006) argues it may also reflect a local view that the threshold is too high. As a result, there may be uneven implementation of guidance, leading to inequitable patient access to new treatments. General perceptions of the HTA process itself can facilitate or hinder implementation. If local providers including prescribing clinicians and other stakeholders view the process as political, ad hoc, or lacking sufficient transparency, they may be more likely to disregard or disagree with decisions or guidance.

To help facilitate the implementation of decisions and the adoption of cost-effective treatments, national and local authorities employ a variety of strategies, ranging from provision of financial planning tools; additional funding to cover the adoption of new technologies; and information dissemination (Sorenson et al., 2008). Sweden uses a network of experts to assist decision makers in understanding guidance and adopting recommended technologies into clinical practice. NICE in the UK takes a similar approach via the use of an in-house Implementation Directorate to ensure that dissemination activities are targeted to local NHS and government representatives as well as the wider community. It assesses and reports on the level of compliance with guidance across the NHS using a variety of data sources on prescribing and practice patterns, examining utilization trends in relation to the expected level consistent with NICE guidance. Reaching local practitioners may be especially important given their role in

the diffusion of technologies – a survey of HTA initiatives in Europe concluded that clinicians frequently fail to change their practice in line with HTA-based recommendations (Banta and Oortwijn, 2000).

Whilst budget holders may have an incentive to resist adoption, financial incentives can be created to reward the use of cost-effective treatments. Several jurisdictions, namely Denmark, Germany, and England, have introduced regulatory levers to make decisions or guidance legally binding, with the latter also using financial incentives through “pay for performance” schemes linked to the uptake of NICE guidance and standards via NHS Quality and Outcomes Frameworks (Drummond et al., 2009a).

Implementation can be further assisted by formal links between the producers and users of HTA. Collaboration between review bodies and local decision makers can foster knowledge and expertise regarding economic evaluation and provide an avenue for local health authorities to lend their perspective. This can assist in developing adoption policies that adequately account for variations in local circumstances. A complementary approach may be for a standard CEA model to be developed, but with flexibility for different decision makers to customize the results by using inputs and evidence relevant to their setting (Drummond et al., 2009a). This approach has been used by US pharmaceutical benefit managers and formularies.

Reassessment after a technology has been used in practice is also an important mechanism to facilitate effective implementation and appropriate technology use. It helps ensure assessments are up-to-date with changes in a technology and the availability of cost-effective products. Several countries, such as France and the UK, have a structured process, conducting re-evaluation at fixed or variable interval (e.g., every 3 to 5 years), while other jurisdictions initiative subsequent reviews if new characteristics of the product emerge or if new or better clinical and/or economic evidence becomes available.

## 7.5 Disinvestment from treatments that are not cost-effective

While these various approaches encourage implementation of decisions and adoption of cost-effective technologies, they may have unintended or unforeseen consequences. For instance, mandates to implement national guidance may steer health care systems towards a suboptimal focus on funding only those interventions that undergo assessment and directing resources away from other, potentially higher value priorities or investments (Birch and Gafni, 2007; Stevens and Milne, 2004). Connected to this issue is the fact that HTA processes often focus on new technologies, giving insufficient attention to existing treatments that may be potentially inefficient or used inappropriately (Drummond et al., 2009a). However, given increasingly limited resources and growing emphasis on value for money, several review bodies, such as NICE and the TLV, are implementing “disinvestment”<sup>20</sup> programmes or strategies.

---

<sup>20</sup> Disinvestment is an explicit process of taking resources from one service in order to use them for other purposes of better value. Rather than a sole focus on allocating new resources, it focuses on eliminating existing ‘waste’ in

The notion of disinvestment makes obvious conceptual sense to ensure efficient resource allocation. In practice, however, removing or limiting currently available services, even if cost-ineffective, raises challenges. There is likely to be opposition from clinicians, interest groups, and patients if existing technologies, services or facilities are no longer available. Local citizens may give a higher value to services at risk, than to technologies they do not yet have, even if the latter deliver more health care than the former (O'Brien et al, 2002; Willan et al, 2001). Many disinvestment initiatives do not get beyond the conceptual stage.

---

the system. However, these technologies may be effective and hence valued by patients and clinicians. They are, however, poor value for money.

## 8. Handling evidence uncertainty in decision making.

### 8.1 The analysis and reporting of uncertainty with sensitivity analysis

The standard approach for allowing for uncertainty in economic evaluations is to conduct a sensitivity analysis. The values of key parameters are varied in order to see whether this greatly affects the result of the study. The most common form is “one-way sensitivity analysis”, where the parameters are varied one at a time. However, “probabilistic sensitivity analysis” (PSA) is increasingly used, although there has been debate about its value alongside the use of patient level data (Griffin et al., 2006). In PSA, probability distributions are specified for all the parameters in the analysis and Monte Carlo simulations are then run in order to estimate the mean and dispersion of the incremental costs and effects. An advantage is that PSA gives an estimate of the overall uncertainty surrounding the estimate of incremental cost-effectiveness of the treatment concerned.

If decision makers are using threshold values of cost-effectiveness, then decision uncertainty can be characterized by estimating the probability that a drug is cost-effective using a range of values of willingness to pay for health gain in a “cost-effectiveness acceptability curve” (CEAC) (Barton et al., 2008). However, in order to assess the treatment option with the maximum expected net benefit (using the threshold willingness to pay for health gain to generate a monetary value) it is also necessary to compare treatment options in a “cost-effectiveness acceptability frontier” (CEAF) (Fenwick et al., 2001). Using a cost-disutility plane has also been proposed as a more helpful framework for analysing and presenting uncertainty (Eckermann et al., 2008). For a review of modeling uncertainty for decision makers with CEA see Briggs et al (2006). McIntosh (2006) has shown how similar approaches can be used to represent uncertainty to decision makers in a WTP or SPDCE analyses. As we note in section 10, the challenge is to represent uncertainty in a way that uses all of the information available and is clearly understood by decision makers.

### 8.2 Value of information and coverage with evidence development

Decisions by payers about the adoption of health technologies are almost always made under uncertainty and on the basis of limited information. In some cases, payers may seek additional research to be undertaken. The use of PSA to characterize uncertainty enables analysts to estimate the “value of information” (VoI) in terms of reducing a decision maker’s uncertainty and therefore potential to make the wrong decision. Wrong decisions have costs, which can include poorer health care for the patients being treated and, indirectly, poorer health care for other patients, if money from a limited health budget is spent on treatments of poor value. In a VoI analysis, the costs of undertaking additional research can be compared with its value to the decision maker (Claxton et al., 2002).

Health care payers therefore have three options (Eckermann and Willan, 2007):

1. “Adopt now” (i.e., use the health technology in question with no additional evidence collection). This is appropriate if both manufacturers and health care payers are confident that the expected clinical and economic value will be realized in real-life or that it is not feasible for practical, ethical or cost (relative to benefit) reasons to collect additional evidence;
2. “Decline” to adopt the health technology in question because it does not accept the expected effectiveness or cost-effectiveness as good value or it is concerned about the degree of uncertainty. Manufacturers have the option of reapplying with more evidence or a different price; or,
3. Adopt the health technology, but make this decision conditional on the collection of further additional evidence. This is usually termed “coverage with evidence development” (CED) which we discuss below.

In making these decisions, the cost and value of additional research should be an important factor. Declining to use a product until more information is available has a cost of delay as well as a cost of undertaking the research. Adopting a drug may make additional research more difficult (although data could be collected in another jurisdiction if it is likely to be transferable) and there may be costs of reversing decisions if subsequent evidence suggests a drug in use is not cost-effective (Eckermann and Willan, 2007).

The use of CED is increasing. There has been a surge of interest in the use of “risk sharing,” a form of CED that involves an agreement between a payer and a pharmaceutical, device, or diagnostic manufacturer, where the price level and/or revenue received is related to the future performance of the product in either a research or real-world environment (Towse and Garrison, 2010; Breckenridge and Walley, 2008). In particular, there is an agreement about a program of data collection to reduce uncertainty about the expected cost-effectiveness of the drug (or device or diagnostic), and the price and/or revenue is linked to the outcome of this program of data collection. This may be prospective or retrospective. Evidence to date has been mixed. In the UK, for example, the MS risk sharing scheme has attracted criticism (Boggild et al., 2009; McCabe et al., 2010).

We can distinguish “risk sharing” from a different type of CED arrangement where payers insist on observational or RCT post-launch data collection when listing the product, but there is *no* pre-agreement as to how to use this data. Renegotiation occurs with new evidence. Several challenges can arise. The evidence collected by the manufacturer may not meet the payers’ requirements (although this could be avoided if evidence requirements were pre-agreed). Indeed, the manufacturer may decide not to collect the information at all in the expectation that the decision to reimburse will not be changed. There may be no pre-agreement on how to interpret the evidence and translate it into a revised price. In most European countries (the UK is an exception) evidence of better effects could not lead to a price increase, whereas in the US, it could. The decision may be restricted to a binary one to adopt or reject – albeit with

improved information. Uncertainty around any pay-offs to additional research will reduce the incentive for the manufacturer to collect the information.

Finally, we can note the potential use of Vol approaches to the challenge of implementing decisions (see Section 7.4). In the same way that the Vol approach can look at the costs and benefits of investment in research to reduce uncertainty about a drug's costs and effects, so it can be used to assess the value versus cost of investing in ways to increase the implementation of the decision. Decision makers can also see whether there is more to gain from investing in research to reduce uncertainty or in getting the decision implemented given the level of uncertainty (Fenwick et al., 2008, Hoomans et al., 2009).

## 9. Are decision makers getting what they want in an efficient manner?

Given the resources required to undertake HTAs and CEAs, it is important that these techniques are being used in a way that generates useful information for decision makers in a cost-effective manner. This has three components. First, the right priorities need to be set for undertaking studies. Second, the costs of studies need to reflect the expected benefits. Thirdly, decision makers need to be able to interpret and use studies appropriately.

### 9.1 Setting priorities for undertaking studies

Regarding priorities for studies, it is unlikely that resources exist to enable every new and existing technology to be investigated. Different jurisdictions use slightly different criteria for determining priorities for CEAs and HTAs (Sorenson et al., 2007, 2008). These normally include criteria such as the importance of the topic in terms of impact on health or health care costs, and the likelihood that the study will reduce the uncertainty surrounding the decision. In the Netherlands, a process exists for determining the priority for undertaking a CEA of new drugs. If the manufacturer accepts that the drug will be “clustered” under the reference pricing scheme, with drugs of similar efficacy and cost, then no CEA will be requested. However, if the manufacturer claims that the product has certain advantages, thereby justifying a premium price, the authorities will request a CEA.

### 9.2 The costs of studies need to reflect the expected benefits

Weinstein (2006) and Buxton (2006) both argue that there are tensions between investing the time and effort to ensure higher quality economic analyses, which are of more potential value, and the risk of reducing their *actual* value because decision makers do not understand the methods and therefore do not trust the results, or the resulting improvement is not worth the cost. In effect, the analytical techniques may lack face validity for decision makers. If so, Weinstein concluded, they are unlikely to be influential and change behaviour. Buxton also outlines the need for more analyses of a wider range of technologies within the resource constraints on appraisals (financial and/or the availability of time and of skilled researchers). Buxton’s concern was that there was an imbalance between the very high quality but very limited quantity of economic evaluations commissioned by NICE in the UK and the lack of evidence of any kind on the cost-effectiveness of most technologies used by the UK NHS.

If increased analytical complexity adds to study costs but leads to better information then it is potentially a worthwhile investment. Whether it is or not depends on both the additional value of the analysis and the opportunity cost of using additional researcher resources. An example, referred to in Section 8.1 above, is criticism of NICE’s insistence on PSA. NICE’s view is that PSA is the best approach to characterizing parameter uncertainty and that practical problems should not compromise this principle. However, NICE has accepted studies that do not have a full PSA.

Several actions might ensure an efficient trade-off of study costs (which to a large part reflect analytical complexity and the opportunity cost of HTA resource constraints) against benefits in terms of decision maker understanding (Towse and Buxton, 2006) including the:

- development of a more contingent appraisal process, where the level of analytical effort is more closely related to the nature and expected impact of tackling the decision-problem;
- use of Value of Information (VoI) methods to determine how much analysis to do in particular cases and which technologies to appraise, if the concerns around face validity for decision makers we discuss above can be addressed; and,
- use of risk-sharing schemes, as discussed above, involving additional data collection with payment contingent on actual patient outcomes. This can allow adoption of technologies despite high levels of uncertainty remaining.

### 9.3 Decision maker ability to interpret and use studies

If we are to obtain value for money from investments in HTA and CEA, decision makers need to interpret and use studies where these are available. As we note in section 9.2., unnecessary study complexity can limit their use if decision makers are unable to interpret them. Another major factor is the lack of institutional infrastructure. Health systems need suitably informed and supported decision-making bodies if they are to use studies in pricing and/or reimbursement decisions. Progress has been made in this regard, with new institutions appearing in several countries. (Sorenson et al., 2008, Drummond et al., 2009a).

The other main route for improving the interpretation and use of CEAs and HTAs would be to make progress in methods for increasing the transferability of studies, as discussed in section 7 above. First, there should be more discussion of the differences in requirements across jurisdictions and whether these are justifiable. Second, there should be more use of methods for adapting studies performed elsewhere for local use, either through the statistical analysis of individual patient data, or through decision analytic modeling, as recommended by Drummond et al. (2009b).



## 10. Conclusions

Theory and practice for measuring the value of pharmaceutical innovations can be summarized as follows:

1. Cost-effectiveness analysis (CEA) has developed as part of a broader approach to the evaluation of health care treatments and programmes, known as health technology assessment (HTA). In the US the term HTA is not widely used, but comparative effectiveness research (CER) is closely linked to HTA. The routine use of CEA rather than cost-benefit analysis (CBA) makes health care different from other areas of public policy.
2. The use of CEA to review the cost effectiveness of pharmaceuticals in their various indications is consistent with principles of efficient resource allocation in health care. Regulating pharmaceutical prices indirectly in this way is the most efficient form of regulation, both theoretically and in practice, if done well (and that of course is the challenge).
3. The quality-adjusted life year (QALY) is the most widely-used measure of health gain in CEAs. It is important that CEAs incorporate a preference-based outcome measure. The challenges of using QALYs partly reflect the experience of using CEA in decision making. Work should be, and is, being done to make the QALY more robust as an accepted measure of the value of a health outcome. Given the theoretical and practical problems with QALYs, alternatives such as willingness to pay estimates and stated-preference discrete choice experiments also need to be explored.
4. The key parameters in CEAs are the assessments of the clinical effectiveness of the health care treatments and programs being evaluated, and the value of the resources they consume. Clinical trials provide important estimates of treatment effect, but usually need to be supplemented by other data (e.g., from observational studies) and synthesized in a decision analytic model in order to make a comprehensive, and potentially generalizable, estimate of clinical and cost-effectiveness. Synthesizing clinical data from different sources in order to improve knowledge of likely effectiveness in a particular setting remains a methodological and practical challenge.
5. Progress could be made in the harmonization of economic evaluation methods and requirements, so as to increase the relevance of CEAs from one jurisdiction to another. A major issue, however, is whether estimates are transferable from one location to another. Often this is not the case. Adjustments need to be made, or local data generated. Payers and manufacturers are only beginning to understand the implications of this issue.
6. In assessing health care programs, it is likely that attributes in addition to health gain are valued. These may include the broader impact of improved health on the economy (taking a societal perspective) and fairness, or equity, in health care provision. In making difficult decisions on the allocation of health care resources, it is important to use a transparent approach for weighting the various attributes of health care

programs. There is a need for more structured approaches to decision making, including the use of MCDA.

7. In budget constrained health care systems it may be helpful to determine a “cost-effectiveness threshold” of willingness to pay, reflecting the opportunity cost of the resources consumed by treatments and programs at the margin. However, determining the level of such a threshold is not a simple task. It may only make sense in health care systems with a “hard” budget constraint. Elsewhere, levels of investment are more appropriately driven by individuals or society’s willingness to pay for improved health. The choice of threshold conveys preferences for innovation. Dynamic efficiency requires the social value of a medicine be reflected in the returns to innovation. In addition there has been debate as to whether thresholds can or should take account of the complexity of incremental gain within a therapy area where R&D externalities (spillover effects) can be important in achieving health gain over time.
8. No matter how well CEAs are conducted, uncertainty always will exist about the estimates. Analytical approaches, such as one-way and probabilistic sensitivity analysis, help in characterizing this uncertainty. The use of “value of information” analysis can help the decision maker determine whether investments in further research might be worthwhile. One approach is to offer “coverage with evidence development” (CED) for drugs and other new health technologies, whereby the new medicine or device is granted market access on condition that further research into its costs and effectiveness is carried out. This can be combined with the use of “risk-sharing” arrangements to adjust price depending on research outcomes. CED could reduce drug development time taken to get to market, especially if linked to conditional licensing approval and early dialogue between licensing agencies, HTA agencies and companies as to what data should be collected in the development process.
9. As with all uses of scarce resources, it is important that the use of CEA itself delivers value for money to decision makers. Deciding which technologies to assess is an investment decision about the use of CEA resources. CEAs need to meet decision makers’ needs without being overly complex if this adds to cost but not to value. It is also important that the mechanisms exist to implement the findings of CEA. Without implementation, investment in CEA cannot deliver a return.

## References

- Appleby, J. et al. (2009) Searching for cost effectiveness thresholds in the NHS. *Health Policy*. 91(3), 239-45.
- Arrow, K. et al. (1993) Report of the NOAA Panel on Contingent Valuation. *Federal Register*. 58, 4601–4614.
- Audit Commission. (2005) *Managing the financial implications of NICE Guidance*. London: Audit Commission.
- Baker, R. et al. (2010) Weighting and valuing quality adjusted life years using stated preference methods: Preliminary results from the social value of a QALY project. *Health Technology Assessment*. 214(27), 1-162.
- Baltussen, R. and Niessen, L. (2006) Priority setting of health interventions: The need for multi-criteria decision analysis. *Cost Effectiveness and Resource Allocation*. 4(14).
- Banta, D. and Oortwijn, W. (2000) Health technology assessment and health care in the European Union. *International Journal of Technology Assessment in Health Care*. 16, 626-35.
- Barbieri, M. et al. (2005) Variability of cost-effectiveness estimates for pharmaceuticals in Western Europe: Lessons for inferring generalizability. *Value in Health*. 8(1), 10-23.
- Barbieri, M. et al. (2010) What do international pharmacoeconomic guidelines say about economic data transferability? *Value in Health*. 13, 102-37.
- Barton, G.R, Briggs, A.H. and Fenwick, E.A.L. (2008) Optimal cost-effectiveness decisions: The role of the Cost-Effectiveness Acceptability Curve (CEAC), the Cost-Effectiveness Acceptability Frontier (CEAF), and the Expected Value of Perfect Information (EVPI). *Value in Health*. 1(5), 86-897.
- Birch, S. and Gafni, A. (2007) Economists' dream or nightmare? Maximizing health gains from available resources using the NICE guidelines. *Health Economics Policy and Law*. 2, 193-202.
- Birch, S. and Gafni, A. (2006) Information created to evade reality (ICER): Things we should not look to for answers. *PharmacoEconomics*. 24, 1121-1131.
- Boggild, M., Palace, J., Barton, P. et al.(2009) Multiple sclerosis risk sharing scheme: Two year results of clinical cohort study with historical comparator. *British Medical Journal*. 339, b4677.
- Breckenridge, A. and Walley, T. (2008) Risk sharing and payment by results. *Clinical Pharmacology and Therapeutics*. 83(5), 666-7.

Briggs, A.H., Sculpher, M.J. and Claxton, K. (2006) *Decision modelling for health economic evaluation*. New York: Oxford University Press.

Brouwer, W.B.F. et al. (2008) Welfarism vs. extra-welfarism. *Journal of Health Economics*. 27, 325-338.

Buxton, M. (2006) Economic evaluation and decision making in the UK. *Pharmacoeconomics*. 24, 1133-1142.

Caro, J.J. et al. (2010) The efficiency frontier approach to economic evaluation of health-care interventions. *Health Economics*. 19(10), 1128-31.

Claxton, K. (2007) OFT, VBP: QED? *Health Economics*. 16, 545-558.

Claxton, K. et al. (2010) *Appropriate perspectives for health care decisions*. Working Paper. York: Centre for Health Economics, University of York.

Claxton, K. et al. (2009) *The value of innovation*. Report by the Decision Support Unit. London: National Institute for Health and Clinical Excellence.

Claxton, K., Sculpher, M. and Drummond, M. (2002) A rational framework for decision making by the National Institute for Clinical Excellence. *Lancet*. 360, 711-715.

Cookson, R. (2005) QALYs and the capability approach. *Health Economics*. 14, 817-829.

Cross, J. and Garrison, L. (2008) *Challenges and opportunities for improving benefit-risk assessment of pharmaceuticals from an economic perspective*. London: Office of Health Economics.

Culyer, A.J. (2009) *Deliberative processes*. London: Office of Health Economics.

Culyer, A.J. (1989) The normative economics of health care finance and provision. *Oxford Review of Economic Policy*. 5(1), 34-58.

Culyer, A.J. (1992) The morality of efficiency in health care -- some uncomfortable implications. *Health Economics*. 1, 7-18.

Culyer, A.J. (1997) The principal objective of the NHS ought to be to maximise the aggregate improvement in the health status of the whole community. *British Medical Journal*. 314, 667-669, 1997. (Reprinted with changes in New, B. ed. (1997) *Rationing: Talk and action*. London: King's Fund and British Medical Journal. 95-100.)

Culyer, A.J. (2008) Resource allocation in health care: Alan Williams' decision maker, the authority and Pareto. In Mason, A. and Towse, A. eds. *The Ideas and Influence of Alan Williams: Be reasonable – do it my way!* Oxford: Radcliffe Publishing. 57-74.

Culyer, A.J. and Lomas, J. (2006) Deliberative processes and evidence-informed decision-making in health care – do they work and how might we know? *Evidence and Policy*. 2, 357-371.

Culyer, A. J. et al. (2007) Searching for a threshold, not setting one: The role of the National Institute for Health and Clinical Excellence. *Journal of Health Services Research and Policy*. 12, 56-58.

Culyer, A.J. and Wagstaff, A. (1993) QALYs versus HYE. *Journal of Health Economics*. 11, 311-323.

Daniels, N. and Sabin, J.E. (2008) Accountability for reasonable priority setting. *British Medical Journal*. 337, a1850.

Daniels, N. and Sabin, J.E. (2002) *Setting limits fairly: Can we learn to share medical resources*. New York: Oxford University Press.

Danzon, P., Towse, A. and Mestre-Ferrandiz, J. (2011) *Value based differential pricing*. NBER working paper. Cambridge, MA: National Bureau for Economics Research. *Forthcoming*

Dintsios, C-M. and Gerber, A. (2010) Some essential clarifications: IQWiG comments on two critiques of the efficiency frontier approach. *Health Economics*. 19(10), 1139-1141.

Department of Communities and Local Government. (DLTR) (2009) *Multi-criteria analysis: A manual*. London: Department of Communities and Local Government. Available at: <http://www.communities.gov.uk/documents/corporate/pdf/1132618.pdf> [Accessed 1 March 2011].

Department of Health (UK). (2009) *The 2009 Pharmaceutical Price Regulation Scheme*. London: HMSO. Available at [http://www.dh.gov.uk/dr\\_consum\\_dh/groups/dh\\_digitalassets/documents/digitalasset/dh\\_098498.pdf](http://www.dh.gov.uk/dr_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_098498.pdf) [Accessed 1 March 2011].

Department of Health (UK). (2010) A new value-based approach to the pricing of branded medicines - a consultation. London: HMSO. Available at [http://www.dh.gov.uk/en/Consultations/Liveconsultations/DH\\_122760](http://www.dh.gov.uk/en/Consultations/Liveconsultations/DH_122760) [Accessed 1 March 2011].

Devlin, N. et al. (2011) *The influence of cost effectiveness and other factors on NICE decisions*. Paper presented at the Health Economics Research Group. University of York. 5-7 January.

- Devlin, N. and Parkin, D. (2004) Does NICE have a cost effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Economics*. 13(5), 437-452.
- Devlin, N. and Sussex, J. (2011) *Incorporating multiple criteria in HTA: Methods and processes*. London: Office of Health Economics.
- De Witt, G.A., Busschbach, J.J.V., and de Charro, F. (2009) Sensitivity and perspective in the valuation of health status: Whose values count? *Health Economics*. 9,109-126.
- Diener, A., O'Brien, B. and Gafni, A. (1998) Health care contingent valuation studies: A review and classification of the literature. *Health Economics*. 7(4), 313-326.
- Dolan, P. (2001) Utilitarianism and the measurement and aggregation of quality-adjusted life years. *Health Care Analysis*. 9, 65-76.
- Dolan, P. and Cookson, R. (1998) *Measuring preferences over the distribution of health benefits*. York: Centre for Health Economics, University of York.
- Dolan, P. and Kahneman, D. (2008) Interpretations of utility and their implications for the valuation of health. *The Economic Journal*. 118(525), 215–34.
- Dolan, P. et al. (2005) QALY maximisation and people's preferences: A methodological review of the literature. *Health Economics*. 14, 197–208.
- Donaldson, C., Mason, H. and Shackley, P. (2006) Contingent valuation in health care. In Jones, A. ed. *Companion to health economics*. Cheltenham: Edward Elgar Publishing.
- Dowie, J. (2008) The future of HTA is MCDA. Version 25. *Knol*. Available at <http://knol.google.com/k/jack-dowie/the-future-of-hta-is-mcda/2rrcx2o0qipsa/14>.
- Drummond, M.F. et al. (2007) Assessing the economic challenges posed by orphan drugs. *International Journal of Technology Assessment in Health Care*. 23(1), 36-42.
- Drummond, M.F. et al. (2009a) Evidence and values: Requirements for public reimbursement of drugs for rare diseases. *Canadian Journal of Clinical Pharmacology*. 2(16), 273-e281.
- Drummond, M.F. et al. (2008) Key principles for the improved conduct of health technology assessments for resource allocation decisions. *International Journal of Technology Assessment in Health Care*. 24(3), 244-258.
- Drummond, M.F. et al. (2005) *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press.
- Drummond, M.F. ed. (2009) Moving the QALY forward: Building a pragmatic road. *Value in Health*. 12(S1).

- Drummond, M.F. et al. (2010) Reimbursement of pharmaceuticals: Reference pricing versus health technology assessment. *The European Journal of Health Economics*. 12(3), 263-271.
- Drummond, M.F. et al. (2009b) Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force Report. *Value in Health*. 12(4), 409-418.
- Drummond, M.F., Ferguson, B., and Weatherly, H. (2008) Economic evaluation of health interventions. *British Medical Journal*. 337, a1204.
- Drummond, M.F., Jönsson, B., and Rutten, F. (1997) The role of economic evaluation in the pricing and reimbursement of medicines. *Health Policy*. 40, 199-215.
- Drummond, M.F. and Rutten, F. (2008) New guidelines for economic evaluation in Germany and the United Kingdom. Are we any closer to developing international standards? *OHE Briefing* (46). London: Office of Health Economics.
- Drummond, M.F. and Sorenson, C. (2009) Use of pharmacoeconomics in drug reimbursement in Australia, Canada and the United Kingdom. What can we learn from international experience? In Arnold, R .G. ed. *Pharmacoeconomics: From theory to practice*. Boca Raton: Taylor and Francis Group.
- Eckermann, S., Briggs, A. and Willan, A. (2008) Health technology assessment in the cost-disutility plane. *Medical Decision Making*. 28, 172-81.
- Eckermann, S. and Willan, A.R. (2007) Expected value of information and decision making in HTA. *Health Economics*. 16, 195-209
- Fanshel, S. and Bush, J.W. (1970) A health status index and its application to health services outcomes. *Operations Research*. 18, 1021-66.
- Fenwick, E., Claxton, K. and Schulpher, M. (2001) Representing uncertainty: The role of cost-effectiveness acceptability curves. *Health Economics*. 10, 779-87
- Fenwick, E., Claxton, K. and Schulpher, M. (2008) The value of implementation and the value of information: Combined and uneven development. *Medical Decision Making*. 28, 21-32.
- Fox-Rushby, J. (2002) *Disability Adjusted Life Years (DALYs) for decision making?* London: Office of Health Economics.
- Fryback, D.G. et al. (2007) Norms for six generic health-related quality-of-life indexes from the national health measurement study. *Medical Care*. 45, 1162-70.

Garber, A. and Phelps, C. (1997) Economic foundations of cost-effectiveness analysis. *Journal of Health Economics*. 16, 1–31.

Garber, A. and Sox, H. (2010) The role of costs in comparative effectiveness research. *Health Affairs*. 29(10), 1805-1811.

Garrison, L. and Towse, A. (2003) The drug budget silo mentality in Europe: An overview. *Value in Health*. 6(s1), S1-S9.

Gold, M., Siegel, J., Russell, L. B. and Weinstein, M. (1996) *Cost-effectiveness in health and medicine*. New York: Oxford University Press.

Griffin, S. et al. (2006) Probabilistic analysis and computationally expensive models: Necessary and required? *Value in Health*. 9(4), 244– 252.

Hill, S. et al. (2003) *Technology appraisal programme of the National Institute for Clinical Excellence*. Copenhagen: WHO Regional Office for Europe.

Hoomans, T. et al. (2009) Value of information and value of implementation: Application of an analytic framework to inform resource allocation decisions in metastatic hormone-refractory prostate cancer. *Value in Health*. 12(2), 315-324.

Hutton, J. et al. (2006) Framework for describing and classifying decision-making systems using technology assessment to determine the reimbursement of health technologies (fourth hurdle systems). *International Journal of Technology Assessment in Health Care*. 21, 10-18.

Insinga, R.P. and Fryback, D.G. (2003) Understanding differences between self-ratings and population ratings for health in the EuroQOL. *Quality of Life Research*. 12(6), 611-619.

International Network of Agencies for Health Technology Assessment (INAHTA). (2010) Available at <http://www.inahta.org/HTA> [Accessed 8 March 2011].

Jena, A.B. and Philipson, T.J. (2008) Cost effectiveness analysis and innovation. *Journal of Health Economics*. 27, 1224-1236

Jena, A.B. and Philipson, T.J. (2009) *Endogenous cost-effectiveness analysis in health care technology adoption*. NBER Working Paper. Cambridge, MA: National Bureau of Economic Research.

Jönsson, B. (2008) IQWiG: An opportunity lost? *The European Journal of Health Economics*. 9(3), 205-207.

Jönsson, B. (2011) Relative effectiveness and the European pharmaceutical market. *European Journal of Health Economics*. 12, 97-102.



Jönsson, B. (2009) Ten arguments for a societal perspective in the economic evaluation of medical innovations. *The European Journal of Health Economics*. 10(4), 357-359.

Jönsson, E. (2002) Development of health technology assessment in Europe. *International Journal of Technology Assessment in Health Care*. 18, 171–183.

Jönsson, E. and Banta, H.D. (1999) Management of health technologies: An international view. *British Medical Journal*. 319, 1293.

Kahneman, D. (2009) A different approach to health state valuation. *Value in Health*. 12, S16-S17.

Kahneman, D. (2007) *QALYs versus experience: A perspective from experimental economics*. London: Office of Health Economics.

Klarman, H.E., Francis, J. and Rosenthal, G. (1968) Cost-effectiveness analysis applied to the treatment of chronic renal disease. *Medical Care*. 6, 48-54.

Lakdawalla, D. and Sood, N. (2009) Innovation and the welfare effects of public drug insurance. *Journal of Public Economics*. 93, 541-8.

Kneis, S. et al. (2009) The transferability of economic evaluations: Testing the model of Welte. *Value in Health*. 12(5), 730–738.

Lakdawalla, D. and Sood, N. (2005) *Insurance and innovation in health care markets*. NBER Working Paper 11602. Cambridge, MA: National Bureau of Economic Research.

Lawrence, D.P., Costa, B.E. and Carlos, A. (2007) Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. *Annals of Operations research*. 154(1), 51-68.

Lipscomb, J. et al. (2009) Retaining, and enhancing, the QALY. *Value in Health*. 12(S1), S18-S26.

Luce, B. et al. (2010) EBM, HTA, and CER: Clearing the confusion. *The Milbank Quarterly*. 88(2), 256-276.

McCabe, C. et al. (2010) Continuing the multiple sclerosis risk sharing scheme is unjustified. *British Medical Journal*. 340, c1786.

McCabe, M., Claxton, K. and Culyer, A.J. (2008) The NICE cost-effectiveness threshold. What it is and what that means. *Pharmacoeconomics*. 26(9), 733-744.

McCabe, C., Claxton, K., and Tsuchiya, A. (2005) Orphan drugs and the NHS: Should we value rarity? *British Medical Journal*. 331, 1016-1019.

- McGuire, A. (2001) Theoretical concepts in the economic evaluation of health care. In Drummond, M. and McGuire, A. eds. *Economic evaluation in health care*. Oxford: Oxford University Press. 1-21.
- McIntosh, E. (2006) Using discrete choice experiments within a cost-benefit analysis framework. *PharmacoEconomics*. 24(9), 855-68.
- Martin, S., Rice, N. and Smith, P.C. (2007) *The link between health care spending and health outcomes: Evidence from English programme budgeting data*. London: The Health Foundation.
- Mason, A. and Towse, A. eds. (2007) *The ideas and influence of Alan Williams. Be reasonable -- do it my way!* New York: Oxford Radcliffe Publishing.
- Mason, H., Jones-Lee, M., and Donaldson, C. (2008) Modelling the monetary value of a QALY: A new approach based on UK data. *Health Economics*. 18(8), 933-950.
- Maynard, A.K. (1991) Developing the health care market. *Economic Journal*. 101, 1277-1286.
- Mehrez, A. and Gafni, A. (1989) Quality-adjusted life years, utility theory and healthy-years equivalents. *Medical Decision Making*. 9, 142-149.
- Meltzer, D. (1997) Accounting for future costs in medical cost-effectiveness analysis. *Journal of Health Economics*. 16(1), 33-64.
- Mishan, E.J. (1971) Evaluation of life and limb: A theoretical approach. *Journal of Political Economy*. 79, 687-706.
- Mitchell, R.C. and Carson, R.T. (1989) *Using surveys to value public goods. The contingent valuation method*. Washington, D.C.: Resources for the Future.
- Murray, C. (1994) Quantifying the burden of disease: The technical basis for Disability Adjusted Life Years. *Bulletin of the World Health Organisation*. 72, 429-445.
- National Institute for Health and Clinical Excellence (NICE). (2009) Appraising life-extending, end of life treatments. London: National Institute for Health and Clinical Excellence. Available at <http://www.nice.org.uk/aboutnice/howwework/devnicetech/endoflifetreatments.jsp?domedia=1&mid=88ACDAE5-19B9-E0B5-D422589714A8EC6D> [Accessed March 2011].
- National Institute for Health and Clinical Excellence (NICE). (2008) Guide to the methods of technology appraisal. London: National Institute for Health and Clinical Excellence.
- Neumann, P.J. (2007) Budget impact analyses get some respect. *Value in Health*. 10(5), 324-325.

- Neumann, P.J. (2004) *Using cost-effectiveness analysis to improve health care opportunities and barriers*. New York: Oxford University Press.
- Neumann, P.J. and Weinstein, M.C. (2010) Legislating against use of cost-effectiveness information. *New England Journal of Medicine*. 363(16), 1495-1497.
- Nord, E. (1995) The person trade-off approach to valuing health care programs. *Medical Decision Making*. 15, 201–8.
- Nord, E. (1999) *Cost-value analysis in health care*. New York: Cambridge University Press.
- O'Brien, B.J. et al. (2002) Is there a kink in consumers' threshold value for cost-effectiveness in health care? *Health Economics*. 11(2), 175-80.
- OECD. (2005) *Health technologies and decision-making*. Paris: Organisation for Economic Co-operation and Development.
- Office of Fair Trading. (2007) *The Pharmaceutical Price Regulation Scheme: An OFT market study*. London: Office of Fair Trading.
- PPACA. (2010) Patient Protection and Affordable Care Act. *U.S. Public Law 111-148*.
- Rawlins, M.D. et al. (2010) Pharmacoeconomics: NICEs approach to decision making. *British Journal of Clinical Pharmacology*. 70, 346-349.
- Rawlins, M.D. and Culyer, A.J. (2004) National Institute of Clinical Excellence and its value judgments. *British Medical Journal*. 329, 224-227.
- Richardson, J. (2007) *Critique and some recent contributions to the theory of cost utility analysis*. Working Paper 77. Melbourne: Centre for Health Program Evaluation.
- Schlender, M. (2008) The use of cost-effectiveness by the National Institute for Health and Clinical Excellence (NICE): No(t yet an) exemplar of a deliberative process. *Journal of Medical Ethics*. 34, 534-539.
- Sculpher, M. and Claxton, K. (2010) Sins of omission and obfuscation: IQWiG's guidelines on economic evaluation methods. *Health Economics*. 19(10), 1132-1136
- Sculpher, M.J. et al. (2004) Generalisability in economic evaluation studies in healthcare: A review and case studies. *Health Technology Assessment*. 8(49).
- Sen, A. (2002) Health: Perception versus observation. *British Medical Journal*. 324: 860–1.
- Sen, A. (1987a) The standard of living: Lecture I, concepts and critiques. *In* Hawthorn, G. ed.

*The standard of living*. Cambridge: Cambridge University Press. 1-19.

Sen, A. (1987b) The standard of living: Lecture II, lives and capabilities. In Hawthorn, G. ed. *The standard of living*. Cambridge: Cambridge University Press. 20-38.

Sendi, P.P. and Briggs, A.H. (2001) Affordability and cost-effectiveness: Decision-making on the cost-effectiveness plane. *Health Economics*. 10, 675-680.

Shah, K.K. (2009) Severity of illness and priority setting in healthcare: A review of the literature. *Health Policy*. 93(2-3), 77-84.

Sheldon, T.A. et al. (2004) What's the evidence that NICE guidance has been implemented? Results from a national evaluation using time series analysis, audit of patients' notes, and interviews. *British Medical Journal*. 329, 999-1003.

Sorenson, C. et al.. (2007) *National Institute for Health and Clinical Excellence (NICE). How does it work and what are the implications for the U.S.?* Arlington, VA: National Pharmaceutical Council.

Sorenson, C., Drummond, M. and Kanavos, P. (2008) *Ensuring value for money in health care: The role of health technology assessment in the European Union*. Copenhagen: WHO Regional Office for Europe.

Stevens, A. and Milne, R. (2004). Health technology assessment in England and Wales. *International Journal of Technology Assessment in Health Care*. 20, 11-24.

Sugden, R. (2008) Citizens, consumers and clients: Alan Williams and the political economy of cost-benefit analysis. In Mason, A. and Towse, A. eds. *The ideas and influence of Alan Williams: Be reasonable -- do it my way*. London: Office of Health Economics and Radcliffe Publishing. 5-18.

Sugden, R. (2005) Coping with preference anomalies in cost-benefit analysis: A market simulation approach. *Environmental and Resource Economics*. 32,129-60.

Sugden, R. and Williams, A.H. (1979). *The principles of practical cost-benefit analysis*. Oxford: Oxford University Press.

Sutton, A., Ades, A.E., Cooper, N. and Abrams, K. (2008) Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics*. 26, 753-767.

Tan-Torres, T. et al. eds. (2003). *WHO guide to cost-effectiveness analysis*. Geneva: World Health Organisation.

The SVQ Research Team. (2008) *Weighting and valuing quality adjusted life years: Preliminary results from the Social Value of a QALY Project*. Report for the UK National Institute for Health Research Methodology Programme.

Torrance, G.W., Thomas, W.H. and Sackett, D.L. (1972) A utility maximization model for evaluation of health care programs. *Health Services Research*. 7, 118–33.

Towse, A. (2007) If it ain't broke, don't price fix it: The OFT and the PPRS. *Health Economics*. 16, 653-665.

Towse, A. and Buxton, M. (2006) Three challenges to achieving better analysis for better decisions: Generalizability, complexity and thresholds. London: Office of Health Economics.

Towse, A. and Garrison, L.P. (2010) Can't get no satisfaction? Will pay for performance help? Toward an economic framework for understanding performance-based risk-sharing agreements for innovative medical products. *PharmacoEconomics*. 28(2), 93-102.

Trueman, P. et al. (2009) The feasibility of harmonizing health technology assessments across jurisdictions: A case study of drug eluting stents. *International Journal of Technology Assessment in Health Care*. 25, 455-462.

Ubel, P., Richardson, J. and Pinto-Prades, J.L. (1999) Life-saving treatments and disabilities. Are all QALYs created equal? *International Journal of Technology Assessment in Health Care*. 15, 738–48.

Weinstein, M. (2006) Recent developments in decision-analytic modeling for economic evaluation. *PharmacoEconomics*. 24(11), 1043-1053.

Weinstein, M.C. and Manning, W.G. (1997) Theoretical issues in cost-effectiveness analysis. *Journal of Health Economics*. 16, 121-128.

Weinstein, M.C. and Stason, W.B. (1977) Foundations of cost-effectiveness analysis for health and medical practices. *New England Journal of Medicine*. 296, 716–721.

Brouwer, W.B.F. and Rutten, F. (2010) The efficiency frontier approach to economic evaluation: Will it help German policy making? *Health Economics*. 19(10), 1128-1131.

Willan, A., O'Brien, B. and Leyva, R.A. (2001) Cost-effectiveness analysis when the WTA is greater than the WTP. *Statistics in Medicine*. 20(21), 3251–3259.

Williams, A. (1997) Intergenerational equity: An exploration of the fair innings argument. *Health Economics*. 6, 117-132.

Williams, A.H. and Cookson, R.A. (2006) Equity-efficiency trade-offs in health technology assessment. *International Journal of Technology Assessment in Health Care*. 22(1), 1-9.