November 2011

OHE Research
Office of Health Economics

seminar briefing no10

WILLINGNESS TO PAY AND PUBLICLY FUNDED HEALTH CARE: CONTRADICTIONS IN TERMS?

Cam Donaldson, Yunus Chair in Social Business and Health Glasgow Caledonian University
Senior Investigator, National Institute of Health Research

CONTENTS

Introduction 1
A brief history of willingness to pay (WTP) 2
WTP for a collectively financed health care good using patients' values: the case of clinical close substitutes 4
WTP for a collectively-financed health care good using public values 6
WTP for national level decision making 8
The fifth defence of WTP 11
Final Remarks 13
Acknowledgements 13
References 14
Appendix 17

Introduction

This paper is about the seemingly contradictory pursuit of "valuing intangibles". More specifically, a major benefit of most health interventions is health itself, or, at least, an improvement in patient or population health. There may be other aspects of health care, of course, that are beneficial to members of society and about which they have preferences, such as where services are located. These health and other benefits often are portrayed as intangible although we would all agree that they have value. Given that health care is a scarce commodity, in the sense of there never being enough resources to meet all claims on the budget, decisions about what (and what not) to spend money on cannot be avoided. This means that values for health and other benefits in different contexts are continuously implied. Thus, beyond recognising that intangibles merely have value, it becomes important to put a number on them because magnitudes of the aforementioned intangibles, and patients' and the public's views of them, will vary from one context to another.

The questions posed in this paper are whether more explicit monetary valuation of benefits, through elicitation of patient and public willingness to pay (WTP) values for different interventions, are not only useful, but also feasible and defensible in a health care system, such as the UK's National Health Service (NHS), that is funded largely from public sources and aimed at allocating resources on the basis of need. In essence, it is a paper presented "in defence of willingness to pay".

The briefing is structured as follows:

- Section 2 provides a brief introduction to the WTP concept, and, through discussing its use in decision making, presents the first and second defence of the method.
- Section 3 analyses the first level of health care decision making at which WTP may be of use, that of aiding clinical dilemmas of the sort often examined in the context of randomised trials. In economics jargon, examination of preference for different ways of treating the same group of patients can be thought of as valuing "close substitutes". This leads to the third defence of WTP, against Reinhardt's famous punch-in-the-nose scenario which questions the basic theory behind the use of WTP values in decision making, and to the fourth defence, against the charge that WTP cannot be used in NHS-type

systems due to its association with ability to pay.

- Section 4 focuses on the second level of decision making where WTP can be used, that of priority setting across health care programmes. Here, no specific defence of WTP is outlined, but it is an important area of research which can be defended in the same ways as in the preceding and forthcoming sections.
- Section 5 examines the third decision-making dilemma in health care, which relates to the use of WTP in decisions at the national level (addressing the question "can we use WTP to estimate the monetary value to be attached to health, or more specifically, quality adjusted life years [QALYs]?").
- This last set of work has led to much controversy about the usefulness of WTP-based surveys of members of the public for valuing QALYs. Therefore, in Section 6, a response to the three main points raised in criticism represents the fifth main defence of WTP presented, before offering some concluding thoughts in Section 7.

A brief history of willingness to pay (WTP)

The term "willingness to pay" (WTP) generally is accredited to a French engineer Jules Dupuit, who used the concept when he was trying to assess whether it was worthwhile building another bridge over the River Seine in Paris:

"Political economy has to take, as the measure of utility of an object, the maximum sacrifice which each consumer would be willing to make in order to acquire the object . . . the only real utility is that which people are willing to pay for" J. Dupuit (1844).

Dupuit's observation was based on the notion of scarcity of resources that underpins economic theory. Every time we use resources we give up the opportunity of using them in alternative ways. The idea is that an economic technique of benefit assessment must involve the notion of sacrifice. More precisely, and for any individual, the maximum sacrifice that a person is prepared to make to attain the object of concern represents the value s/he attaches to that object. A societal value is then reached through aggregating such values across the relevant population – which may also require netting out of values of those against any particular policy or good.

The idea that value entails some kind of trade off also applies to the main methods underlying the dominant metric for measurement of benefits in economic evaluation of health care, the QALY. The main methods underlying QALY valuation are standard gamble (SG) and time trade-off (TTO). In the former, individuals are asked to sacrifice certainty for the more uncertain prospect of better health, whilst, with TTO, the sacrifice is more obviously made between time at the end of life and health gain now. In discrete choice experiments, sacrifices are made through respondents being asked to trade levels of different attributes against each other. With WTP, it is money that is sacrificed. Money has the advantage of being flexible in that it can be interpreted as representing all other goods on which a given respondent's WTP amount could have been spent. It is also important to note at this point that SG, TTO and WTP are all forms of "stated preference" approach, whereby, through the use of survey methodology, valuations are drawn directly from individuals largely through asking them to respond to hypothetical scenarios. (Although not the main point of this paper, it is fair to say that there is no doubt that the hypothetical nature of WTP elicitation processes is an important challenge. However, as summarised in Donaldson, et al. (2011), and in our first defence of WTP, it is also fair to say that the literature is split as to whether or not this is a problem, with leaders in the field claiming that adverse results tend to result from sub-standard instrument design (Carson, et al., 1996). Related to this is the issue of which payment vehicle to use in WTP elicitation. These can be: open-ended, whereby a question is posed of a respondent about their WTP for a good, with no guidance provided as to what the answer might be; closed-ended (or referendum), whereby a value is presented to a respondent to which they would answer "yes" or "no" as to whether s/he would be willing to pay, an approach which tends to require larger sample sizes than others; and payment scales, where a range of values is presented to each respondent for them to select which is their maximum. Simple illustrations of each of these are provided in the Appendix, but there are several variations, such as bidding games and, one that we have tended to use much more frequently in interview-based studies, where a set of cards is presented to respondents in random order and they decide for each card whether or not they would pay – this seems to combine the best of all of the above approaches.)

With these basics of hypothetical values and payment vehicles out of the way, let us return to Jules Dupuit. Consistent with his approach, the focus of this paper is on using WTP to assess strength of preference for a health intervention for use in an economic evaluation of whether to provide such an intervention though public funding. Such studies, as reported here, are not about using WTP to assess how much people would be prepared to tolerate in user charges or other forms of private payment. In terms of empirical applications, it took more than a century for Dupuit's observations to be tested. Davis (1963) published the first empirical study using WTP in the public policy context, using the method to place a value on outdoor recreation. Jones-Lee (1974) used WTP to value risk reductions in the

area of safety and transport. In the first study in the health arena, Acton (1976) used WTP to value risk reductions for heart attacks. This was followed by another pause, at least in health, until the early 1990s, when more WTP studies began to be published (Donaldson, 1990; Johannesson, et al., 1991).

It is fair to say that this break was likely due to the popularity of QALY valuation methods at the time. It took some time to start questioning the QALY fundamentals and identifying the important elements that were missing, coinciding with a re-emergence of WTP. For example, two key areas of weakness of the QALY approach identified in the literature have been:

- Sensitivity, as the QALY does not perform well in capturing small changes in health that might be picked up by more condition-specific measures (Donaldson, et al. ,1988);
- Process, as the QALY does not factor into its calculus non-health process-type attributes over which patients may have preferences (Donaldson and Shackley, 1997).

Several classic articles critique QALYs, particularly: Loomes and McKenzie (1989), which explores, amongst other things, the issue of the linear and proportional nature of QALYs, whereby a gain of 0.1 on the usual 0-1 scale (where 0=death and 1=full health) is assumed to have the same value (in QALY terms) no matter what is starting point; Bleichrodt (1995), who criticises QALYs on the basis of the separability assumption, whereby quality of life is valued independently of life years before the two are combined; and the whole body of work by Birch and Gafni which, much to the author's puzzlement, seems to have been largely ignored in applied health economic evaluation, despite offering useful criticism of the QALY's lack of a basis in utility theory (Gafni and Birch, 2000). In sum, and in our second defence of WTP, one could argue that WTP has a sound basis in welfare economics, in part because it is less restrictive in allowing more factors to be included in the evaluation of preferences and in how these factors are combined with each other. With the WTP approach, when you present people with a scenario about a particular type of health care that you are asking them to pay for, they can bring to that calculation anything they wish. QALYs restrict respondents to valuing only health-related quality of life and only in terms of risk (SG) or time (TTO). Despite admirable attempts to do so, in this author's view, proponents of QALYs have struggled to come up with an alternative theoretical basis for them (see Culyer, 1989; Birch and Donaldson, 2003).

A framework for the use of WTP in decision making

The way in which WTP is used in decision making critically depends on the nature of the good to be valued and on whose values are to be used. Figure 1, from (Donaldson and Shackley, 2000), attempts to illustrate this point.

Figure 1: *Nature of the good, and the values to be used.*

		Values elicited from...	
		...patients	...public
Nature of funding	Private	(1)	(2)
	Collective	(3)	(4)

In the case of a privately funded good, this will simply be an "add on" to what is currently insured, and the decision can be based on whether the aggregated WTP values (of either patients or the public) is greater than the cost of adding the relevant procedure to the list of what is covered, i.e., if benefits (B) > costs (C), it should be funded. When the good is collectively financed, the situation is more complex.

With a cash limit to consider, the opportunity cost of provision or expansion will be another procedure that is not funded.

Considering the other relevant element of whose values should be used, perhaps simplistically, patients might be the best people to ask if a decision has to be made about the funding of two different ways of treating a particular group of people. Public values can be more appropriate for broader resource allocation problems.

Figure 1 shows four main scenarios that then follow from these considerations, with their associated decision rules being as follows:

(1) Private financing using patients' values: if B>C the intervention evaluated should be funded
(2) Private financing using public values: Again, if B>C the intervention evaluated should be funded
(3) Collective financing using patient's values: Although it may seem natural to ask patients about the value attached to alternative therapies for their condition, the results of such studies are not straightforward in terms of their use. Obviously if WTP for one therapy is greater than that for its rival, and cost for the former is also less, this could be argued to represent an unambiguous situation where the more beneficial (according to WTP) therapy would be

recommended. However, if the former is also more costly, then its potential provision cannot be considered in isolation from other uses of these resources, which, in a cash-limited system, will mean that other types of patients may lose out. This leads to scenario (4).

(4) Financing a collectively funded good using public values: Shackley and Donaldson (2000) characterise this scenario by describing several projects competing for health care funds at the margin. Some of these projects may be candidates for scaling back and some for expansion. The key information the decision maker then requires is that on the public's WTP and on the cost for each of these alternatives. The set of interventions chosen would be that which maximises total benefit (total WTP) for the marginal funds at stake.

Of course, it is possible to mix these scenarios. For example, one might take more of an "insurance-based" approach to WTP in a publicly-funded system and thus ask the public to consider its WTP to add services to what is already covered. Indeed, this approach was outlined in the Canadian context many years ago by O'Brien and Gafni (1996).

The following two sections of the paper, however, will focus on development of the WTP approach in the context of scenarios (3) and (4). These scenarios can be mapped onto what might be called two decision makers' dilemmas. With scenario (3), the dilemma is essentially a clinical one in the sense of deciding which intervention is better for a given group of patients. With scenario (4), the WTP framework is suitable for analysing "across-programme" dilemmas of the sort that a health authority might face in deciding how to best allocate resources to meet the needs of its designated population. The classic study in this respect is that by Olsen and Donaldson (1998) on WTP for helicopters, hearts and hips.

Since the work of Shackley and Donaldson (2000), a third level of decision making has leapt to prominence. This is at the national level, involving health technology assessment agencies (such as the National Institute for Health and Clinical Excellence, or NICE) making one-off decisions about whether to recommend a specific intervention for funding by the health care system. Given that many such agencies use cost per QALY gained as the basis for such recommendations, this raises the question of whether WTP can be used to value such QALY gains in monetary terms and even aid judgement about what the "threshold" value of a QALY should be. Therefore, section five addresses WTP research that has followed from this development.

WTP for a collectively financed health care good using patients' values: the case of clinical close substitutes

In line with the more clinical notion of the randomised trial, many economic evaluations take place in the context of addressing the question of which type of care to provide for a given group of patients. Indeed, the resurgence of WTP in health in the early 1990s started largely within this context. As will be seen, this initially led to some problems with the method.

WTP began to be used within the context of randomised trials because many such studies were struggling to show differences between groups in terms of the generic quality of life measures used. It was thought that WTP might reveal differences in patient preferences that these other measures could not. Despite this logic, a basic challenge is that, when one thinks about it, WTP questions can be asked in at least three ways in the context of a trial.

The first approach pursued was to ask patients in a trial for their maximum WTP for the treatment they receive. Those having treatment A would give a WTP for A, V(A), whilst those having treatment B would provide a value V(B). Several early studies of WTP in health, however, found that this approach did not work well in discriminating between treatments, and it was speculated that this was because respondents are unaware of the fact that there is an alternative on offer and so tended to provide a value for the care they have had relative to "no treatment". Thus, patients in each arm of the trial were effectively valuing the same thing; treatment versus no treatment. Hence the lack of ability of WTP studies of this type to discriminate between treatments evaluated (Donaldson, et al., 1995; Johannesson and Fagerberg, 1992; Ryan, et al., 1997).

Therefore, in my own work with several colleagues, we then proceeded to devise elicitation approaches that tried to focus the respondent on the differences between alternatives, on the basis that this was what the original randomised trial was trying to determine. Inadvertently, this began to lead us to approaches more in line with economic theory – which, of course, is perhaps where we should have started, rather than from a more clinical perspective!

The approaches we devised were:

1. First, "WTP for each", where we still elicited a value for each alternative, but this time with each respondent estimating V(A) and V(B); and
2. Second, and followed by, the "marginal approach", where each respondent provided a value for their

preferred over their less-preferred option, V(A-B) or V(B-A).

Practically, each of these approaches requires the WTP study to be conducted with a group of patients similar to those involved in the trial, but after the trial has been completed. This is to avoid any bias that might arise from a respondent having experienced one or other of the interventions being evaluated. Going back to theoretical considerations, these approaches each involve making respondents aware of the choice between a good and a close substitute. This, in a sense, recognises the importance of "reference points" used in prospect theory (Kahneman and Tversky, 1979). Furthermore, each approach can involve collection of data on respondents' preferences; indeed the marginal approach requires this. A person's WTP for his/her preferred option can then be compared with that for his/her less-preferred option, making it easier to test (at the individual respondent level) the validity of the approach, i.e., quite simply is the direction of WTP responses in line with the simple statement of preference ordering?

Using this basic test, the WTP-for-each approach ran into problems because it was shown to result in significant preference reversals, i.e., lack of congruence between simple preferences and the magnitudes of WTP responses. Most likely, this is because respondents compare the cost of alternatives on offer and base their WTP on that rather than on strength of preference for each (Donaldson, et al., 1997a).

Quite quickly, we arrived at the marginal approach. Again, in the context of a randomised trial, respondents would be asked to choose between the status quo (current care) and an alternative (experimental care), and then asked for their maximum WTP to have their preferred option instead of that less preferred. The aim is to clarify to respondents that it is their additional valuation of their preferred option that is required, regardless of whether the preferred option is more or less costly than their less-preferred option. The method also fits more closely with Kaldor-Hicks theory (Hicks, 1939; Kaldor, 1939). Of course, strictly, the Kaldor-Hicks criterion would state that if those individuals gaining from a proposed programme could compensate the losers and still remain better off, the proposed programme should go ahead, whether or not the compensation can be paid. Although the last part of this statement is problematic (see below, in this section), a simple interpretation of this rule is that a change should be made if the expressed value to gainers (say, those who prefer an experimental therapy) is greater than the value of losses associated with those who lose out from the change (say, those who prefer the status quo, or current care), assuming costs are not under consideration.

Studies using the marginal approach have resulted in WTP values that discriminate between options on offer, and, by including losers as well as gainers from a given policy change (or retention of the status quo), have shown that net benefits would have been overestimated if respondents had been asked to value the experimental treatment only (Donaldson, et al., 1997b; Gibb, et al., 1998). Furthermore, in some cases, it has been shown that the (apparently) less costly alternative received a higher mean WTP than the alternative option, and in none of the studies did respondents indicate cost as a reason underlying their stated WTP value. Thus, it seems that the marginal approach is promising and further studies of its validity continue.

Two related defences of WTP are required at this point. Continuing the sequence from above, our third defence of WTP stems from the famous "punch-in-the-nose" example of Reinhardt (1992), used in his critique of the Kaldor-Hicks criterion. The basics of his example are that me punching you on the nose represents an overall improvement in social welfare if I can compensate you for doing this whilst I still remain better off, even although I do not actually compensate you. Health economists who are not too keen on WTP seem to draw much mirth from this example, but it too has its problems. First, it does not deal with a situation where resources are at stake; the key being how welfare gains and losses from alternative uses of resources balance out. Second, it is not clear where the punch-in-the-nose example leads us in terms of information needs. Surely, in a situation of resource scarcity it is important to know who gains and who loses from alternative uses of resources. This information can only be revealed from elicitation of preferences. This would be stage 1 in an evaluation process based on Kaldor-Hicks, as such preferences reveal who are the gainers and who are the losers. What is done beyond point may be more open to question, but this would seem to me to be a key piece of information. From here, whether we go down a WTP or QALY route, they each have the same normative basis in the compensation principle of Kaldor-Hicks (Birch and Donaldson, 2003). Although rarely recognised in cost-utility analysis, QALYs gained are being compared with QALYs lost all the time and QALYs, too, suffer from the same distributional challenges as WTP (Donaldson, et al., 2002).

Following this last point, the fourth defence of WTP is with respect to distributional issues. The marginal approach provides a useful illustration of why such issues may not be as problematic as one might first think when using WTP. First, knowing people's preferences, the researcher can investigate whether people in a higher income group tend to choose one option more frequently than those in lower income groups. Second, the researcher can also examine

whether WTP values from those in higher-income groups distort the overall value given for one or other option. A scheme along these lines for at least diagnosing whether ability to pay is a problem was outlined by Donaldson (1999a) who also showed how sensitivity tests can be used to estimate a break-even distributional weight at which the final decision on which option to choose would alter. This work showed that, to alter initial (unweighted) results, such weights have to be set at a draconian level, such that the preference of higher-income groups would hardly count at all. This is in line with the classic work of Harberger (1978):

"The lessons from these examples is clear: when distributional weights are used together with weighting functions of the type most commonly employed in writings on the subject, the result is to open the door to projects and programs whose degree of inefficiency by more traditional (unweighted) cost-benefit methods would (I feel confident) be unacceptable to the vast majority of economists and informed public" (Harberger, 1978, p. S113).

Of course, it is not necessary for studies of patients' values always to be tied to a randomised trial. One alternative is to ask patients to value the health care good in question, whilst another is to ask only for a valuation of a characteristic of the good (such as the health gain it might provide, or information if it is a diagnostic tool), with this value being fed into a broader cost-benefit analysis, a distinction drawn by Currie, et al. (2002).

Finally, it is important to note that this section has been used to describe how WTP values can be employed in what are largely ex post situations. Reiterating a point touched on above, if a WTP study indicates that significant benefits would be gained by implementing a new option, but the new option would also lead to increased costs, then in most health care systems which operate with a fixed budget, the opportunity cost of implementation will manifest itself as benefits forgone by some other group of patients. There is an argument that values between such competing options are better elicited in an ex ante situation where members of the community consider alternatives based on some idea of risk to themselves and/or other members of the community needing such care. Examples of this ex ante approach are provided in the following two sections.

WTP for a collectively-financed health care good using public values

The second level of decision-making dilemma arises in organisations such as health authorities. Usually, such authorities are allocated a fixed amount of resources per annum from which they have to choose the combination of services to provide to meet the needs of their designated populations. Such health authorities start with a given mix of resources at any point in time. If any wishes to change this mix within the context of a fixed budget, logic would dictate that the authority may have to decide on which services are candidates for expansion and contraction, and evaluate them accordingly, before deciding whether to maintain the current mix or proceed with some expansions at the expense of some contractions. To make this decision, the authority would need to weigh up the costs and benefits of each service competing at the margin to be in or out of the budget. Originally, the main research question for health economists thus became whether we could obtain WTP-based community values on the benefits generated by the competing candidates.

This question was addressed through an extensive collaboration over many years with Jan Abel Olsen from the University of Tromso. Because of its remote location, in Tromso, and other places in northern Norway, a helicopter ambulance service is provided. During the early 1990s, when this service was being evaluated, some sort of monetary WTP-type measure was thought to be a useful approach of assessing its benefits. This led to our first study: the Northern Norway (or "helicopters, hearts and hips") study (Olsen and Donaldson, 1998).

Initially, the aim was to obtain a monetary valuation from the population at large for the helicopter ambulance for use in a cost benefit analysis (CBA). However, if such a service were to use resources from a fixed public sector pot, an opportunity cost would arise through not having the resources available for other services. Therefore, each of 150 respondents was asked to rank the ambulance service against providing 80 more elective heart operations and 250 more elective hip operations, each described in terms of who would get treated (e.g., men aged 50-60 years) and expected outcomes in the presence and absence of the intervention. These alternatives were chosen because existing evidence made them relatively easy to describe and because they permitted us to compare values for an emergency service (the helicopter) with two non-emergency services, one which was life-extending and the other which was quality-of-life-enhancing only. The respondents were told that these options were competing for funding from their community. After ranking them, respondents were asked for their WTP in extra taxation per annum for each option. In Norwegian Kroner (NK), WTP for a helicopter was NK316 on average, whilst the corresponding values for the heart and hips programmes were NK306 and NK232 respectively. Strict interpretation of these results in WTP

terms would mean that the helicopter and hearts were valued higher than the hip intervention.

As the first of its kind, this study had several flaws, the most obvious being that, with hips always asked about last, respondents may have reached some kind of budget constraint that deflated the hips WTP values. This is despite respondents having been told that the options were competing with each other for limited funds and that successive WTP amounts expressed were not additive. Furthermore, building on the use of simple preference data, it was possible to observe that initial rankings of the three programmes given by respondents were not matched by rankings implied by WTP values (Olsen, 1997).

Further developments in the field took place at the same time as the Northern Norway study. For example, suggestions had been made for taking an "insurance-based" approach to WTP questioning, whereby respondents would get information on their own (or their families') probabilities of needing care (O'Brien and Gafni, 1996). This contrasted with the "community-based" approach in Olsen and Donaldson (1998) where enhancements in programmes were described as being for the community (rather than using individual probabilities explicitly) and payments were through taxation, under which substantial numbers of people mentioned altruistic reasons for being willing to pay.

These flaws and parallel developments led to a project which gathered together a wider network of researchers around Europe who were interested in improving the method of WTP in the context of eliciting community values for priority setting. "EuroWill", a project funded by the European Commission, involved surveys in six European countries (Donaldson, 1999b). As well as addressing the issues listed above, the project addressed several other issues that had arisen in the WTP literature during the 1980s and 1990s. Along with other details of the surveys in each country, these are listed in Table 1, taken from Donaldson, et al. (2011). Amongst other things, the work has shown: the importance the

Table 1: *Issues addressed and numbers of responses in each country*

Country (dates of survey)	Areas of care	Issues addressed	Numbers receiving different versions of the questionnaire	Total in each country
Norway (March 1997)	More heart operations More cancer treatments Helicopter ambulance	Insurance versus community-based questions Size of effects	Community-based = 80 Insurance-based = 83 Community-based (2 cancer progs) = 79 Community-based (all progs for less people) = 81	323
Portugal (Oct-Nov 1997)	More heart operations More cancer treatments Improved car ambulance	Numbers treated	Cancer, hearts, ambulance = 104 Cancer, hearts, improved hearts = 103 Cancer, improved hearts, ambulance = 103	310
Denmark (April 1998)	More heart operations More hip operations More cataract operations	Insurance-based versus community-based questions Test–retest	Insurance = 168 Community = 168 Test-retest = 50	386
UK (May-June 1998)	More heart operations More cancer treatments Helicopter ambulance	Payment scale versus closed-ended	Payment scale = 236 Closed-ended = 342	578
France (Oct-Nov 1998)	More heart operations More cancer treatments Helicopter ambulance Reduction in pollution	Process utility Cognitive capacity Test-retest	No process = 100 Neutral process = 104 Positive process = 99 Pollution/no process= 154 Pollution/process = 149 Test-retest = 50	353
Ireland (April 1999)	More heart operations More cancer treatments More community care	Marginal approach Ordering effects	Basic approach = 113 Marginal approach = 121 Different ordering = 101	335
Grand total				2285

Source: Donaldson, et al. (2011)

population puts on community vis-à-vis acute services; the existence of ordering effects, reasons for them and potential solutions; the potential for improving consistency between explicit and implied rankings by using a marginal approach, whereby respondents are asked for their WTP for their least preferred programme and then how much more than that for their more preferred programmes; the difficulty of detecting scope effects (with respect to different sizes of health gain or different numbers of people benefiting); that the closed-ended approach yields significantly higher WTP values than the payment scale; that different values are obtained by providing respondents with different amounts of information about the same option; and that improved econometric techniques are feasible for estimating factors associated with WTP where multiple alternatives are valued by each respondent (O'Shea, et al., 2002; Stewart, et al., 2002; Shackley and Donaldson, 2002; Olsen, et al., 2004; Ryan, et al., 2004; Protière, et al., 2004; Luchini, et al., 2003).

With respect to distributional issues using across-programme WTP values elicited from the general public, the main issue is whether distributions of WTP by income group for each option follow the same pattern (Olsen and Donaldson, 1998).

Note also that values from the public can be elicited in publicly-funded systems, but in slightly different contexts and even using a combination of approaches from this section and the previous one. For example, a variation of the marginal approach has been applied to explore public preferences on fluoridation of water supplies (Shackley and Dixon, 2000) and, thus, challenges with respect to distributional issues could be addressed in the same way as outlined in section 3. Also, as discussed above, in other countries, where the issue might be whether an intervention should be added to the "menu" of items covered, from public or private funds and at an extra cost to payers, WTP questions could be asked about the single intervention being considered, an approach taken by O'Brien, et al. (1998) in the context of provision by a health maintenance organisation in the US. Here, distributional challenges would be dealt with in similar fashion to the approach outlined in the following section.

WTP for national level decision making

The third level of decision making dilemma is that of one-off decisions at national (sometimes regional or provincial or state level in more-federated situations). This has reached prominence in policy and in empirical research due to the creation of national level health technology assessment agencies (Hirth, et al., 2000; Gyrd Hansen, 2003; Byrne, et al., 2005; King, et al., 2005). When assessing particular interventions in terms of health gains against the costs of provision, such agencies must, in effect, put a monetary value on those health gains. In the context of England, where the National Institute for Health and Clinical Excellence (NICE) uses the QALY as its health metric, NICE must decide what value(s) of a QALY to use.

Since the inception of NICE, the threshold value of a QALY has been set at £20-30,000 (Rawlins and Culyer, 2004). Interventions with a cost per QALY above this range are less likely to be recommended by NICE for adoption by the rest of the NHS (Devlin and Parkin, 2004; Clement, et al., 2009). The threshold was based on best guesses of experts at the inception of NICE, and has been subject to criticism since the UK House of Commons Health Committee (2002) review of NICE in 2001-2002. The criticism at that time centred on the lack of an empirical basis for the threshold. More recently, pressure has been placed on NICE to raise the threshold, as exemplified in the case of life-extending drugs for people in the terminal phase of cancer (Lakhani, 2008; Richards, 2008), and to lower it, based on forthcoming fiscal pressures and views of primary care trusts (PCTs) that NICE guidance is not affordable (West, 2009). Indeed, arguments for and against raising the threshold were debated in a recent head-to-head in the BMJ (Towse, 2009; Raftery, 2009).

Partly in response to such criticism, NICE co-funded two research projects in 2004. The author was principal investigator on one of these, the Social Value of a QALY (SVQ) project, a collaboration between the Universities of Newcastle upon Tyne, East Anglia and Aberdeen. Around the time of their publication (Mason, et al, 2009; Baker, et al., 2010), the results of SVQ were used in some quarters to put further upward pressure on the threshold (Lakhani, 2008; Richards, 2008; Towse, 2009; Smith and Porter, 2008). However, use of the results in such reporting was selective and it became important that a summary overview of the project and interpretation of its results was presented by those who had conducted the research. Much of the remainder of this section draws upon that summary (Donaldson, et al., 2011).

The value of a QALY may be sought in different ways. Valuable information has been generated, for example, in analysis of the affordability and cost of generating a QALY at the level of PCTs, given their pre-determined budgets (Martin, et al., 2008; Appleby, et al., 2009). SVQ consisted of three related strands, each based on eliciting values from members of the general public. One of these was concerned with eliciting QALY weights and is not of relevance to this particular paper.

SVQ work on modelling the value of a QALY

The first strand of SVQ involved modelling the monetary value of a QALY from the WTP-based value of preventing a statistical fatality (VPF) that the UK Department for Transport (DfT) and other public sector agencies apply to life-saving projects. This value is derived from asking representative samples of the public about their WTP for safety improvements. These improvements are characterised as reducing the risk of death for any individual by small amounts in the forthcoming time period (e.g., the coming year). Across a population, a small number of actual lives will be saved. In simple terms, dividing aggregate WTP of the population by this small number of lives saved gives us the VPF. With WTP values being elicited from a cross-section of the population, it can be argued that the resulting VPF (or value of a QALY, if that is the focus of interest) is reflective of society's overall budget constraint. An important ethical standpoint is that the resulting "average" value is applied to each member of society regardless of income. Indeed, public sector agencies that employ WTP-based values (such as the DfT and the Health and Safety Executive in the UK) invariably do apply the same value, based on the population average, to all income groups.

A simplified version of the method of transforming the VPF into a value of a QALY is as follows:

> A straightforward way to compute the value of a QALY is to start with the well-established roads VPF for the UK. For example, if we take a representative death avoided as being that of a person aged 35, assume that the VPF is £1.4m (or £1.4x106) and that the person concerned would have lived for another 40 years, a rough calculation of the value of a life year gained by that person would be as follows:
>
> $$V = \frac{£1.4 \times 10^6}{40} = £35{,}000$$
>
> Conveniently, V is close to the value of a QALY espoused by Rawlins and Culyer (2004). However, if one were to assume that not all of the 40 years gained would be spent in full health (especially later years) and a discount rate applied, the denominator would fall, thus raising the value of a QALY above £35,000. For example, if the discount rate was taken to be 3.5% then the annualised sum that would have a discounted present value of £1.4m over 40 years would be £77,300.

However, the value resulting from this would reflect a particular QALY type. By QALY-type, we mean that QALYs can be generated in at least two ways: by adding years to life or by enhancing the quality of remaining life years without extending life. The former can be further subdivided into avoiding immediate threats to life or adding years to the end of life. The procedure outlined above reflects the first of these, although a more-sophisticated approach, still using the VPF also was used to model the value of a QALY arising from life extension as opposed to life saving (Mason, et al., 2009). Similar WTP procedures also were used by DfT to derive the VSI, or value of a serious injury prevented (Department for Transport, 2007), from which a value of a quality-of-life-enhancing QALY can be derived (Mason, et al., 2009), summarised as follows:

> Each serious injury was broken down into 3 phases; in hospital effect (valued at 0.69 or 0.16 on the EQ-5D tariff, depending on severity of injury and generally modelled as lasting for one month), initial after-effects (generally for two months and valued at 0.76) and longer-term effects (for remaining life and valued at either 0.76 or 0.3, again depending on severity). Assuming that any given injury would occur at the mean age of the UK population, with 26 expected remaining QALYS, we calculated an overall total QALY loss for each scenario. We then divided the VSI of £150,000 by the total QALY loss for each scenario and computed a weighted average based on probability of each scenario occurring.

Thus, values for all three QALY types could be explored within the research.

Note that, beyond this, SVQ did not look at the value attached to QALY gains from treating specific diseases. This is due to a more generic rather than disease-specific approach to economic evaluation being the tradition in UK health economics and decision making.

Table 2 gives a typical set of values of a QALY that have arisen from the modelling. It would seem that different "QALY-types" would imply different values. Based on WTP to reduce the risks of life-threatening events, values close to £70,000 per QALY were produced, as compared to values around £35,000 for a life-extending QALY. Estimating gains from improvements in quality of life, with no increase in number of remaining years, produced a lower value of about £10,000 per QALY.

Table 2: *Values of a QALY via alternative calculations from modelling based on VPF and VSI*

Basic modelling approach	Value of a QALY (£)
Life-saving	£70,000
Life-extending	£35,000
Quality-of-life-enhancing	£10,000

Nevertheless, note that the above results imply that the threshold could be raised for life-saving QALYs. One note of caution, however, involves reflecting on what might be meant by a life-saving QALY being valued at seven times that of a quality-of-life-enhancing QALY. Although not calculated like this, it might be reasonable to assume that the "average" utility score across the profiles of serious injuries were, say, 0.6. This would imply that WTP to avoid a fatality would be just over 11 times that for the serious injury (i.e. 70,000/10,000 x 1.0/0.6). This, along with the results in Table 1, therefore presents a hypothesis that requires further testing. (Note that the issue of WTP for QALY types was not explored directly in survey work in SVQ, as was also the case with the issue of end-of-life QALYs, but has been incorporated into subsequent surveys on the "European value of a QALY" (see the EuroVaQ website at http://research.ncl.ac.uk/eurovaq/). Moreover, the VPF itself is just over nine times the value of the VSI. The inference that there is no single value of a QALY is in line with other published views (Brouwer, et al., 2009) and the lowest value is also reflective of earlier published WTP studies looking at the value of a QALY (Gyrd-Hansen, 2003).

SVQ feasibility survey on the valuing of a QALY

The second strand of SVQ assessed the feasibility of obtaining an estimate of the monetary value of a QALY by presenting members of the public with appropriately framed valuation questions in a survey. Example health states are as follows:

> Stomach: 3 months
> Initially you will have severe stomach pains, diarrhoea, vomiting and fever for 7 days, severe enough to interfere with most of your usual activities.
>
> Things then improve, but for up to one year from initial onset you will suffer an episode of stomach discomfort and sickness every couple of weeks, with each episode lasting for 2-3 days. These episodes are not so severe but may interfere with some of your usual activities.
>
> (Half of the respondents were given stomach health state descriptions of 3 months, 12 months and lifetime durations.)
>
> Head: 3 months
> You will have episodes of throbbing pain across the front of your head and you will feel sick and may occasionally be sick. You will feel like you want to lie still in a darkened room.
>
> During the next three months you will suffer an episode of head pain and sickness every couple of weeks, with each episode lasting between 8 hours and two days. These episodes will interfere with many of your usual activities. After three months you return to your current health with no further effects from this illness.
>
> (The other half of the respondents was given head health state descriptions of 3 months, 12 months and lifetime durations.)

An example question to illustrate how changes in quality of life and WTP were estimated and combined was as follows:

> The value of a QALY is derived via a "chaining" procedure. In the initial part of the chain, the respondent is asked about whether s/he would be prepared to pay anything to avoid being in this state, and, if so, what is the maximum amount s/he is willing to pay.
>
> In the second part of the chain, the respondent would be asked a "standard gamble" question involving a choice between two options. In the standard way of deriving a QALY index, one option would leave the respondent in the stomach/head condition for certain for the remainder of his/her life whilst the other option would involve a gamble with varying probabilities of a better or worse outcome. "Better" usually means a return to full health for the rest of one's life, whilst worse is usually characterised as immediate death. Visual procedures are used to guide the respondent through the process, and the index is derived from the point at which the respondent feels it is difficult to choose between the outcome for certain and the gamble.
> Let us assume that, for one respondent, the probability at which s/he finds it difficult to choose between the head condition for certain and taking the gamble is 0.95 and that his/her WTP to avoid a year in the head condition was £1000. Dividing £1000 by 0.05 (which comes from subtracting 0.95 from 1) would give a value of a QALY for that person of £20,000. This can be done across several individuals to arrive at an average value of a QALY for a population.
>
> For either head or stomach conditions, each respondent was asked two WTP questions (to avoid the three-month state and the 12-month state) and three standard gamble questions (3 months for certain vs. a gamble with outcomes of return to current health or 12 months in the state; 12 months for certain vs. a gamble with outcome of return to current health or rest of life in the state; and rest if life for certain vs. gamble with outcomes of current health or immediate death).

In fact slightly more WTP and standard gamble questions were asked of each respondent, but these are not relevant to this paper.

From this, it can be seen that any individual respondent would be faced with a set of WTP and standard gamble questions, the two sets then being combined in different ways to arrive at values of a QALY. Respondents could have been asked time trade-off questions instead of standard gambles, our rationale for the latter simply being that it comes from the same theoretical "stable" as WTP. On the other hand, given that the QALY tariff used by NICE is based on TTO (using 10-year time horizons), it is not necessarily the case that direct comparison between the value of a QALY derived from SVQ and NICE's valuation of a QALY can be made. Also, rather than combine WTP values with a pre-existing tariff (such as that for the EQ-5D quality of life system), we wanted our own respondents to provide health state utility values that could be combined with their own WTP values for purposes of internal consistency. Eliciting WTP from our sample and then combining these with EQ-5D values from a different population also would have been problematic. It is also important to note that the aim of this part of SVQ was to test the feasibility and robustness of the elicitation methods; and, rather than conduct a full-scale national sample survey, the sample was restricted to 400 people, which was not representative of the population.

This survey work suggested that it is feasible to conduct a survey to elicit monetary values for a QALY from a representative sample of the public so long as the procedure is broken down into manageable steps and is carried out on a face-to-face basis by well-trained interviewers. However, it also became apparent that the mean estimates produced by such questions are particularly prone to the influence of "outlier responses" and that great care is therefore required in the selection of central-tendency measures. The most common example of an outlier was that many people were willing to take only very small risks of a more adverse outcome to avoid the stomach and head health states in the standard gamble questions, or were even not willing to gamble at all. As well as such floor effects, respondents may also have a WTP ceiling (or budget constraint), an amount they express whether for a small or large perceived gain. Thus, when WTP values and health state utilities are combined in such circumstances, the implied WTP per QALY for such individuals can be so high as to lead to an implausible population average WTP per QALY across the whole sample. This was indeed the case in SVQ, with the value running into several millions of pounds! This result was consistent with that of Pinto Prades, et al. (2009).

Other ways of managing the data, therefore, are displayed in Table 3. Rather than computing a ratio of WTP/QALY loss for each individual and then taking a mean, the first two calculations take the mean WTP, the mean QALY loss and then compute the ratio. This is done for each of the stomach and head scenarios. The third calculation takes a ratio of medians. So, for example, using median stated WTP to avoid the certainty of a 12-month period of illness, the figures suggest a value for a QALY in the region of £20,000-£40,000.

There is no doubt that SVQ made substantial contributions to the health economics and health policy literatures, but already the research has moved on. In chaining the values derived from standard gamble and WTP questions in the SVQ survey work, there appeared to be a methodological problem in identifying health states that are serious enough to encourage more respondents to trade them (against risks of death and full health) in standard gambles, but that are not so serious that paying for their avoidance is perceived as unaffordable in WTP questions. This problem also led to implausibly high WTP-per-QALY values when these were derived from individual values of a QALY. This has led to a new chaining procedure having been devised and implemented in the "European value of a QALY" project (see the EuroVaQ website at http://research.ncl.ac.uk/eurovaq/). EuroVaQ was a collaboration, funded by the European Commission, covering 10 countries and involving an Internet-based survey of about 40,000 respondents across Europe. As the results are not yet published, they cannot be detailed here. However, the innovation to encourage trading worked to such a degree that values of a QALY could be derived from taking means and medians of individuals' own values of a QALY, and the eventual values derived are not out of line with earlier studies and with those used in current policy. EuroVaQ also examined the issue of the value of QALY types, showing quality-of-life-enhancing QALYs to be valued less than those involving life extensions, although results on end-of-life QALYs were less clear cut.

The fifth defence of WTP

My fifth defence of WTP for a QALY is based on a collection of points that have arisen in what I would call the debate between "searchers" and "surveyors" (Baker, et al., 2011). The surveyors are people like me, who have used survey research methods to elicit WTP-based values of a QALY from members of the general public. The searchers' position is basically that a series of investment and disinvestment decisions (by, say, NHS entities like Primary Care Trusts [PCTs]) can be examined in order to establish the cost per QALY at the margin at which such entities appear to be operating; in a sense, the value of a QALY at the margin can be searched for and, of course, arrived at. If the

Table 3: *Values of a QALY via calculations from survey research*

Using 12-month (mean) WTP value and (mean) standard QALY index:

CONDITION	CALCULATION	VALUE OF QALY
Stomach	1. Mean WTP to avoid 12-month S = £1870 2. Average QALY benefit = 0.104 Divide 1. by 2.	£17,980
Head	1. Mean WTP to avoid 12-month H = £3250 2. Average QALY benefit = 0.144 Divide 1. by 2.	£22,570

Using 3-month (mean) WTP value and (mean) standard QALY index:

CONDITION	CALCULATION	VALUE OF QALY
Stomach	1. Mean WTP to avoid 3-month S = £810 2. Average QALY benefit = 0.026 Divide 1. by 2.	£31,150
Head	1. Mean WTP to avoid 3-month H = £1495 2. Average QALY benefit = 0.036 Divide 1. by 2.	£41,530

Using 12-month (median) WTP value and (median) standard QALY index:

CONDITION	CALCULATION	VALUE OF QALY
Stomach	1. Median WTP to avoid 12-month S = £500 2. Median QALY benefit = 0.025 Divide 1. by 2.	£20,000
Head	1. Median WTP avoid 12-month H = £1000 2. Median QALY benefit = 0.025 Divide 1. by 2.	£40,000

Source: Baker, et al. (2010); Donaldson, et al. (2011)

disinvestments made by PCTs were to have a lower cost per QALY gained than the recommendations made by NICE, this could be taken as evidence that the NICE threshold is too high.

Thus, the main the main criticism of the surveyors by the searchers has been that, as the NHS budget is set through negotiation between the Treasury and the Department of Health, individual WTP values have no relevance for health service resource allocation as they are somehow detached from the budget-setting process, as reflected in the following:

"information about how much an individual or society values improvements in health (i.e. their willingness to pay for a QALY) is not at all relevant to the NICE remit" (Culyer, et al., 2007, p. 57).

In this debate, the WTP approach has even been described as a threat to constitutional jurisdiction of Parliament! The notion here is that survey-based WTP approaches somehow substitute the "direct democracy" of public opinion for the parliamentary process through which the Treasury and Department of Health agree a budget for the NHS, within which PCTs and NICE have to work. It is claimed that "experimental methods" (i.e., WTP studies) cannot capture the opportunity costs more effectively than Parliament as the budget-setting process for the NHS takes account of assessments of the marginal value of extensions of a wide range of public programmes and of the value of purchasing power left in the pockets of consumers (McCabe, et al., 2008).

Given these substantial points, how can a WTP-based approach possibly be defended? The details of the arguments presented in the following paragraphs in this section are contained in Baker, et al. (2011), although some additional points that my co-authors would not allow me to make there have been added!

The notion that WTP-based values of a QALY are detached from the NHS budgetary process is a powerful argument; but it can be countered in three ways. First, NICE itself could be subjected to the same criticism of being detached from the NHS budget. NICE does not have a budget and PCTs constantly complain about decisions being imposed upon them by NICE. Such complaints, if valid, would mean that the current threshold is leading to a misallocation of NHS resources which, ultimately, means we are getting less health for our tax £s spent.

Another anomaly in the searchers' arguments that somehow NICE can search for a threshold, via a series of decisions it has made and will make, is that one would predict that the value to emerge from this process is likely to be around £30,000 per QALY. This is because this is the value at which NICE already operates! The literature abounds with decision makers getting these things wrong in the past. It may well be that £30,000 is the correct number because it was made on the basis of intelligent guesses by experts at NICE's outset. But we can never know if this is the case or whether there is a path dependency towards £30,000 because that is the number that is now "out there". Indeed, the influence of the stated threshold on decisions has already been demonstrated in the literature (Devlin and Parkin, 2004).

The third response is largely theoretical. However, when answering WTP questions in surveys, it is conceivable that the fully-informed respondent would think of the NHS as being at full efficiency and unable to provide more services (or more QALYs) without extra payments being made. If this is the case, expressed WTP amounts would be a reasonable representation of a value of a QALY at the margin for the NHS and may not be far away from what an individual budget-holder, like a PCT, might say is the value (if PCTs used QALYs and if they behaved in an economically rational and QALY-maximising fashion!). Actually, both approaches, searchers' and surveyors', are required, and the Baker, et al. (2011) paper outlines in theory why this is the case, thus offering a reconciliation with respect to this debate.

Finally, on the idea of WTP-based values being a threat to the Constitution, the simple response is to pose the following question: What can be more legitimate than asking those who stand to gain or lose from different allocations of NHS resources – i.e., the public – what are their views about what should count? NICE does this anyway, as the quality-of-life "tariff" that it uses to calculate the cost per QALY of interventions is based on a survey of the general public. Furthermore, if bodies such as NICE wish to go beyond having a single value of a QALY to valuing QALY types, the only legitimate way of doing this (and of avoiding politicians and decision-makers being criticised for the decisions made) is to establish more thoroughly the preferences of the public as regards these different types.

Final Remarks

The main point to arise from the above presentation is that major advancements have been made over the last 20 years in creating a suite of WTP tools that can be used as a decision making aid at different levels of the health care system. Some of these developments have been unique even within the broader areas of economics and contingent valuation as well as health economics alone.

WTP as a method has been subjected to at least as much scrutiny and criticism as any other method of benefit valuation in health economics. I will leave for the reader to decide whether it is robust, in absolute or relative terms, to criticisms arising from the hypothetical nature of most studies, different approaches leading to different results, theoretical critiques (like the punch-in-the-nose example), distributional concerns and the idea of being detached from health care budgets. However, I appreciate the opportunity to have made the case for the defence!

Acknowledgements

The contents of this paper represent 25 years of developing and applying methods to elicit willingness to pay for health and health care goods. During that time, I have been fortunate to work with many individuals, but would like to pay special tribute to Phil Shackley, Jan Abel Olsen and Rachel Baker, with whom I have collaborated over most of this period. I would also like to thank Gillian Currie, Dorte Gyrd Hansen, Graham Loomes, Helen Mason and Angela Robinson with whom I have worked on specific projects. Since venturing into willingness to pay methods in the late 1980s, I have received amazing support and stimulation from Michael Jones-Lee. It was one of the greatest privileges of my career to work with Mike from 2002-2010 whilst I was Health Foundation Chair in Health Economics at Newcastle University.

References

Acton, J.P. (1976) Measuring the monetary value of lifesaving programs. *Law and Contemporary Problems.* 40(4), 46-72.

Appleby, J., Devlin, N., Parkin, D., Chalkidou, K. and Buxton, M. (2009) Searching for cost effectiveness thresholds in the NHS. *Health Policy.* 91(3), 239-245.

Baker, R., Batekman, I., Donaldson, C., Jones-Lee, M., Lancsar, E., Loomes, G., Mason, H., Odejar, M., Pinto Prades, J.L., Robinson, A., Ryan, M., Shackley, P., Smith, R., Sugden, R. and Wildman, J. (2010) Weighting and valuing quality-adjusted life-years using state preference methods: Preliminary results from the social value of a QALY project. *Health Technology Assessment.* 14(27), 1-178.

Baker, R., Chilton, S., Donaldson, C., Jones-Lee, M., Lancsar, E., Mason, H., Metcalf, H., Pennington, M. and Wildman, J. (2011) Searchers versus surveyors in estimating the monetary value of a QALY: Resolving a nasty dilemma for NICE. *Health Economics, Policy and Law.* 6(4), 435-447.

Birch, S. and Donaldson, C. (2003) Valuing the benefits and costs of health care programmes: Where's the 'extra' in extra-welfarism? *Social Science & Medicine.* 56(5), 1121-1133.

Bleichrodt, H. (1991) QALYs and HYEs: Under what conditions are they equivalent? *Journal of Health Economics.* 14(1), 17-37.

Brey, R., Loomes, G. and Pinto-Prades, J.L. (2009) Trying to estimate a monetary value for the QALY. *Journal of Health Economics.* 28(3), 553-562.

Byrne, M.M., O'Malley, K. and Suarez-Almazor, M.E. (2005) Willingness to pay per quality-adjusted life year in a study of knee osteoarthritis. *Medical Decision Making.* 25(6), 655-666.

Carson, R.T., Flores, N.E., Martin, K.M. and Wright, J.L. (1996) Contingent valuation and revealed 80-99. *Land Economics.* 72(1), 80-99.

Clement, F.M., Harris, A., Li, J.J., Yong, K., Lee, K.M. and Manns, B.J. (2009) Using effectiveness and cost-effectiveness to make drug coverage decisions. *Journal of the American Medical Association.* 302(7), 1437-1443.

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

Culyer, A., McCabe, C., Briggs, A., Claxton, K., Buxton, M., Akehurst, R., Sculpher, M. and Brazier, J. (2007) Searching for a threshold, not setting one: The role of the National Institute for Health and Clinical Excellence. *Journal of Health Services Research & Policy.* 12(1), 56-58.

Currie, G., Donaldson, C., O'Brien, B., Stoddart, G.L., Torrance, G.W. and Drummond, M.F. (2002) Willingness to pay for what? A note on alternative definitions of health care programme benefits for contingent valuation studies. *Medical Decision Making.* 22(6), 493-497.

Davis, R.K. (1963) Recreation planning as an economic problem. *Natural Resource Journal.* 3(2), 239-249.

Department for Transport (UK). (2007) *Highways economic note no 1, 2005.* London: Department for Transport.

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.

Donaldson, C. (1990) Willingness to pay for publicly-provided goods: A possible measure of benefit? *Journal of Health Economics.* 9(1), 103-118.

Donaldson, C. (1999a) Valuing the benefits of publicly-provided health care: Does ''ability to pay'' preclude the use of ''willingness to pay''? *Social Science & Medicine.* 49(4), 551-563.

Donaldson, C. (1999b) *Developing the method of 'willingness to pay' for assessment of community preferences for health care.* Final report to Biomed 2 Programme (PL950832) of the European Commission. Aberdeen: University of Aberdeen.

Donaldson, C. and Shackley, P. (1997) Does "process utility" exist? A case study of willingness to pay for laparoscopic cholecystectomy. *Social Science & Medicine.* 44(5), 699-707.

Donaldson, C. and Shackley, P. (2000) Willingness to pay for publicly-financed health care: How should we use the numbers? *Applied Economics.* 32(15), 2015-2021.

Donaldson, C. et al. (2010) *European value of a Quality Adjusted Life Year.* Project FP6. Report for the European Commission. Newcastle-upon-Tyne: Newcastle University. Available at http://research.ncl.ac.uk/eurovaq/EuroVaQ_Final_Publishable_Report_and_Appendices.pdf [Accessed 26 October 2011].

Donaldson, C., Atkinson, A., Bond, J. and Wright, K. (1988) Should QALYs be programme specific? *Journal of Health Economics.* 7(3), 239-257.

Donaldson, C., Baker, R., Mason, H., Jones-Lee, M., Lancsar, E., Wildman, J., Bateman, I., Loomes, G., Robinson, A., Sugden, R., Pinto Prades, J.L., Ryan, M., Shackley, P., Smith, R. (2011) The social value of a QALY: Raising the bar or barring the raise? *BMC Health Services Research.* 11(8). Available at http://www.biomedcentral.com/1472-6963/11/8 [Accessed 7 November 2011].

Donaldson, C., Birch, S., and Gafni, A. (2002) The distribution problem in economic evaluation: Income and the valuation of costs and consequences of health care programmes. *Health Economics.* 11(1), 55-70.

Donaldson, C., Farrar, S., Mapp, T., Walker A. and Macphee, S. (1997b) Assessing community values in health care: Is the 'willingness to pay' method feasible? *Health Care Analysis.* 5(1), 7-29.

Donaldson, C., Mason, H. and Shackley, P. (2011) Contingent valuation in health care. In Jones, A. ed. *The elgar companion to health economics.* Cheltenham: Elgar.

Donaldson, C., Shackley, P. and Abdalla, M. (1997a) Using willingness to pay to value close substitutes: Carrier screening for cystic fibrosis revisited. *Health Economics.* 6(2), 145-159.

Donaldson, C., Shackley, P., Abdalla, M. and Miedzybrodzka, Z. (1995) Willingness to pay for antenatal screening for cystic fibrosis. *Health Economics.* 4(6), 439-452.

Drummond, M.F. and O'Brien, B. (1994) Statistical versus quantitative significance in the socioeconomic evaluation of medicines. *PharmaEconomics.* 5(5), 389-398.

Dupuit, A.J.E. (1844) On the measurement of the utility of public works. Reprinted in Arrow, K.J. and Scitovsky, T. eds. *Readings in welfare economics.* Homewood, IL: Richard D. Irwin.

Gafni, A. and Birch, S. (2000) Preferences for outcomes in economic evaluation: An economic approach to addressing economic problems. *Social Science & Medicine.* 40(6), 767-776.

Gibb, S., Donaldson, C., Henshaw, R. (1998) Assessing strength of preference for abortion method using willingness to pay. *Journal of Advanced Nursing.* 27(1), 30-36.

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

Gyrd-Hansen, D. (2003) Willingness to pay for a QALY. *Health Economics.* 12(12), 1049-1060.

Harberger, A.C. (1978) On the use of distributional weights in social cost-benefit analysis. *Journal of Political Economy.* 86(2), S87-S120.

Hicks, J.R. (1939) The foundations of welfare economics. The Economic Journal. 49(196), 696-712.
Hirth, R.A., Chernew, M.E., Miller, E., Fendrick, M. and Weissert, W.G. (2000) Willingness to pay for a quality-adjusted life year: In search of a standard. *Medical Decision Making.* 20(3), 332-342.

House of Commons Health Committee (UK). (2002) National Institute for Clinical Excellence. *Second report of session 2001-02.* London: Her Majesty's Stationary Office.

Johannesson , M., Jönsson, B. And Borgquist, L. (1991) Willingness to pay for antihypertensive therapy — results of a Swedish pilot study. *Journal of Health Economics.* 10(4), 461-473.

Johannesson, M. and Fagerberg, B. (1992) A health-economic comparison of diet and drug treatment in obese men with mild hypertension. *Journal of Hypertension.* 10(9), 1063-1070.

Jones-Lee, M. (1974) The value of changes in the probability of death or injury. *The Journal of Political Economy.* 82(4), 835-849.

Kahneman, D. and Tversky, A. (1979) Prospect theory: An analysis of decision under risk. *Econometrica.* 47(2), 263-291.

Kaldor, N. (1939) Welfare propositions and interpersonal comparisons of utility. *The Economic Journal.* 49(195), 549-552.

King, J.T., Tsevat, J., Lave, J. and Roberts, M.S. (2005) Willingness to pay for a quality adjusted life year: Implications for societal resource allocation. *Medical Decision Making.* 25(6), 667-677.

Lakhani, N. (2008) The Price of life: It was £20,000. Now an NHS agency recalculates. *Independent on Sunday.* 12 October, 18-19.

Loomes, G. and McKenzie, L. (1989) The use of QALYs in health care decision making. *Social Science & Medicine.* 28(4), 299-308.
Luchini, S., Protière, C. and Moatti, J.P. (2003)

Evaluating several willingness to pay in a single contingent valuation: Application to health care. *Health Economics.* 12(2), 51-64.

Martin, S., Rice, N. and Smith, P.J. (2008) Does health care spending improve health outcomes? Evidence from English programme budgeting data. *Journal of Health Economics.* 27(2), 826-842.

Mason, H., Jones-Lee, M. and Donaldson, C. (2009) Modelling the monetary value of a QALY: A new approach based on UK data. *Health Economics.* 18(8), 933-950.

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

O'Brien, B. and Gafni A. (1996) When do the "dollars" make sense? Toward a conceptual framework for contingent valuation studies in health care. M*edical Decision Making.* 16(3), 288-299.

O'Brien, B.J., Goeree, R., Gafni, A., Torrance, G.W., Pauly, M.V., Erder, H., Rusthoven, J., Weeks, J., Cahill, M., and Lamont, B. (1998) Assessing the value of a new pharmaceutical: A feasibility study of contingent valuation in managed care. Medica*l Care.* 36(3), 370-384.

O'Shea, E., Stewart, J., Donaldson, C. and Shackley, P. (2002) Eliciting preferences for resource allocation for health care. *Economic and Social Review.* 32(3), 217-238.

Olsen, J.A. (1997) Aiding priority setting in health care: is there a role for the contingent valuation method? *Health Economics.* 6(6), 603-612.

Olsen, J. and Donaldson, C. (1998) Helicopters, hearts and hips: Using willingness to pay to set priorities for public sector health care programmes. *Social Science & Medicine.* 46(1), 1-12.

Olsen, J.A., Donaldson, C. and Periera, J. (2004) The insensitivity of 'willingness to pay' to the size of the good: new evidence for health care. *Journal of Economic Psychology.* 25(4), 445-460.

Pinto Prades, J.L., Loomes, G., Brey, R. (2009) Trying to estimate a monetary value for the QALY. *Journal of Health Economics.* 28(3), 553-562.

Protière, C., Donaldson, C., Luchini, S., Moatti, J.P. and Shackley, P. (2004) The impact of information on non-health attributes on willingness to pay for multiple health care programmes. *Social Science & Medicine.* 58(7), 1257-1269.

Raftery, J. (2009) Should NICE's threshold range for the QALY be raised? No. *British Medical Journal.* 338, 268-269.

Rawlins, M. and Culyer, A. (2004) National Institute for Clinical Excellence and its value judgements. *British Medical Journal.* 329, 224-7.

Rheinhardt, U. (1992) Reflections on the meaning of efficiency. Can efficiency be separated from equity? *Yale Policy and Law Review.* 10(20), 302-315.

Richards, M. (2008) *Improving access to medicines for NHS patients.* London: Department of Health.

Ryan, M., Ratcliffe, J., and Tucker, J. (1997) Using willingness to pay to value alternative models of antenatal care. *Social Science & Medicine.* 44(3), 371-380.

Ryan, M., Scott, D. and Donaldson, C. (2004) Valuing health care using willingness to pay: A comparison of the payment card and dichotomous choice methods. *Journal of Health Economics.* 23(2), 237-258.

Shackley, P. and Dixon, S. (2000) Using willingness to pay to elicit public preferences for water fluoridation. *Applied Economics.* 32(6), 777-787.

Shackley, P. and Donaldson, C. (2000) Willingness to pay for publicly financed health care: How should we use the numbers? *Applied Economics.* 32(15), 2015-2021.

Shackley, P. and Donaldson, C. (2002) Using willingness to pay to elicit community preferences for health care: Is an incremental approach the way forward? *Journal of Health Economics.* 21(6), 971-991.

Smith, R. and Porter, A. (2008) NHS ordered to place more value on life. *Daily Telegraph.* 4 November. 1.

Stewart, J., O'Shea, E., Donaldson, C. and Shackley, P. (2002) Do ordering effects matter in willingness to pay studies of health care? *Journal of Health Economics.* 21(4), 585-599.

Towse, A. (2009) Should NICE's threshold range for the QALY be raised? Yes. *British Medical Journal.* 338, 268-269.

APPENDIX

CLOSED-ENDED WTP QUESTION

ANTI-HYPERTENSION TREATMENT

This question concerns how you personally value your treatment against high blood pressure. Since the treatment against high blood pressure claims a lot of health care resources a possible development is that patients in the future will have to pay a larger proportion of the treatment cost, in the form of higher user fees. Assume that this will be the case and that user fees are raised. At present a patient treated for high blood pressure pays on average SEK 350 a year in user fees for drugs and physician visits. Would you choose to continue your current treatment against high blood pressure, if the user fees for the treatment were raised with SEK 2,000* per year?

☐ YES

☐ NO

* The amount was varied between SEK 100 and SEK 10,000 in 15 sub samples

Source: Johannesson, et al. (1991)

CYSTIC FIBROSIS SCREENING STUDY:

For the purposes of the questionnaire we would like you to imagine that you live in a country like the USA where people do have to pay for tests. Because of the way the health service is run in the UK, there is no question of you being asked to actually pay any money.

We are interested in the value you place on having the cystic fibrosis carrier test. One way of doing this is to ask you how much you would theoretically be willing to pay for the test.

Open-ended

1. What is the maximum amount of money you would be prepared to pay for the test? (Please write in the space below).

£ _____

Payment scheme in which respondent gets a scale from which to choose a value

1. What is the maximum amount of money you would be prepared to pay for the test?

Put a ✔ next to the amounts that you are sure you would pay	£0 £1 £2 £4 £6 £8
Put a ✘ next to the amounts that you are sure you would not pay	£10 £12 £16 £20 £30
Put a circle around the maximum amount you would be prepared to pay	£50 £75 £100 £100+

Source: Johannesson, et al. (1991)

About the Office of Health Economics

Founded in 1962, the OHE's terms of reference are to:

- commission and undertake research on the economics of health and health care;
- collect and analyse health and health care data for the UK and other countries;
- disseminate the results of this work and stimulate discussion of them and their policy implications.

The OHE's work is supported by research grants and consultancy revenues from a wide range of UK and international sources.

The research and editorial independence of the OHE is ensured by its Policy Board and Editorial Board.

Further information about the OHE can be found at www.ohe.org

Office of Health Economics
Southside 7th Floor 105 Victoria Street London SW1E 6QT
Telephone: +44 (0)20 7747 8850
Facsimile: +44 (0)20 7747 8851
www.ohe.org

© Office of Health Economics