

Using willingness to pay to elicit community preferences for health care: further evidence from using a ‘marginal approach’

by

Phil Shackley¹ and Cam Donaldson^{1,2}

¹ School of Population and Health Sciences, University of Newcastle, 21 Claremont Place, Newcastle upon Tyne, NE2 4AA, UK

² Business School, University of Newcastle, UK

Paper presented at the Health Economists’ Study Group Meeting, Oxford, 5th – 7th January 2005

Introduction

Recent reviews of the literature on the use of WTP to assess the benefits of health care indicate that the technique is becoming increasingly popular (Diener *et al.*, 1998; Klose 1999; Olsen and Smith, 2001; Shackley, 2004). While the technique has been applied in many different contexts, the majority of applications have focused on eliciting values from patients, where patients can be defined as individuals who are ‘currently diseased’. By this is meant individuals who may be in the process of consuming health care, or are waiting to consume health care (e.g. on a waiting list), or who are immediately eligible to consume health care (e.g. pregnant women offered antenatal screening). When asked to value the health care programme under consideration, patients will be primarily concerned with the value to them (and perhaps also their immediate family) from consuming the health care. Less common are studies which elicit values from members of the public, where the public can be defined as anyone who is not a patient relative to the condition being considered. This definition includes not only individuals who are at risk of contracting a disease (and therefore at risk of consuming the treatment programme) but also individuals who are not necessarily at risk. For example, if a cervical screening programme were being evaluated, the public would include both women and men. When asked to value the health care in question, the public may be concerned with the value to them (and again perhaps their immediate family) from consuming the care, but they may also be more concerned than patients with wider benefits such as option value and externalities (Weisbrod, 1964; Culyer, 1971).

Irrespective of from whom values are elicited, the focus of the majority of studies has been the elicitation of values for alternative interventions for treating a common condition. There are very few examples of studies in which WTP values are elicited for health care programmes to treat disparate conditions. A notable exception is the group of studies arising from the EuroWill project (Olsen *et al.*, 2004a; Olsen *et al.*, 2004b; Protière *et al.*, 2003; Ryan *et al.*, 2004; Shackley and Donaldson, 2002; Stewart *et al.*, 2002). EuroWill was a European Community funded project, the principle aim of which was to investigate the feasibility of eliciting WTP values from members of the public for a range of health care programmes to treat quite different conditions, with a view to using the values to aid decision making in a broader priority setting context (Donaldson, 1999). The essence of the approach was to elicit values for a number of proposed expansions to existing health care programmes.¹ Separate studies were undertaken in six different European countries, with each study addressing one or more methodological issues pertinent to the elicitation of WTP values in this context. For example, the sensitivity of WTP values to changes in the size of the benefit being considered (i.e. scope sensitivity) was investigated in Norway and Portugal (Olsen *et al.*, 2004a), the effect of eliciting WTP values using a payment card compared to an closed-ended approach was the subject of the UK survey (Ryan *et al.*, 2004), the impact of the amount and type of information presented to respondents was investigated in France (Protière *et al.*, 2004), the effect of different question formats on WTP values was the focus of the study undertaken in Denmark (Olsen *et al.*, 2004b), while the effect of presenting options to individuals in different orders was one of the issues investigated in Ireland (Stewart *et al.*, 2002).

One of the more notable results from the EuroWill project was the general finding across surveys that there tended to be a sizeable proportion of respondents for whom the explicit ranking of the three programmes was inconsistent with the ranking implied by their WTP

¹ This emphasis on changes at the margin was adopted so as to reflect the situation facing decision makers in budget constrained publicly-financed health care systems, such as the UK NHS, whereby it is relatively small changes to existing provision which are most often contemplate rather than the introduction or abandonment of entire health care programmes.

values, i.e. there were apparent preference reversals (Olsen, 1997). In response to this, an attempt was made in the Ireland survey to develop a method of asking the WTP questions which was expected to improve consistency between the explicit and implied ranks – the method being termed the marginal approach. The results of this study were inconclusive insofar as while there was no improvement in consistency from using the marginal approach, there was evidence to suggest that this may have been due to design issues rather than the method itself (Shackley and Donaldson, 2002). In light of this, the primary aim of this paper is to investigate further whether the marginal approach has the potential to improve consistency by reporting the results of an empirical study in which the design issues identified in the previous study are addressed.

A secondary aim is to investigate the feasibility of asking respondents to value not only proposed programme expansions, but also proposed reductions to existing programmes. The reason for doing this is to allow for the very real possibility that at any one time, a public sector health care decision maker will be faced with a number of proposals for expansion which cannot be funded from resources which have become available from elsewhere (an implicit assumption of the EuroWill approach, which focused on expansions only). Given fixed budgets, it is likely that the extra resources required to fund any expansions will have to come from existing health care programmes, i.e. reductions in the provision of health care programmes will have to be contemplated. That this variation is at least as likely (if not more likely) to occur than the situation implicitly assumed in the EuroWill surveys is evidence by the increasing interest shown in the development of programme budgeting and marginal analysis (Mitton and Donaldson, 2001). In such circumstances, it is important that preferences are elicited for both the proposed expansions and reductions.

Background

EuroWill

As already stated above, the broad aim of the EuroWill project was to assess and develop the WTP methodology as a tool for measuring strength of preference of members of the public for publicly-funded health care programmes. This was achieved by undertaking interview based surveys in six European countries in which WTP values were elicited for expansions to three disparate health care programmes. The structure of the questionnaire was broadly the same in each country and was as follows (for more details of the EuroWill surveys, see, for example, Shackley and Donaldson, 2002). After a brief introduction by the interviewer, respondents were presented with descriptions of the three health care programmes and asked to rank them in terms of how important they felt they were. Following this, respondents were asked to value the programmes in terms of WTP. The elicitation of WTP values took the form of a partial evaluation, i.e. the evaluation of each programme was conducted independently of the other two programmes. Respondents were told to disregard the other two programmes when stating a WTP value for the third, the implication of this being that after each programme evaluation, respondents' income was restored, i.e. their 'purchasing power' was the same for each programme. Reasons for respondents' answers were elicited. A series of person trade-off questions then followed before the interview was concluded with a series of socio-demographic questions.

The marginal approach

The finding that for significant numbers of respondents in the surveys that the explicit ranking of the programmes did not match the implied ranking from WTP values led to the development of an alternative method of posing the WTP questions – the marginal approach

(Shackley and Donaldson, 2002).² In essence the technique involved taking the lowest ranked programme and eliciting a partial WTP value for this first. The second ranked programme was then presented to respondents who were asked to state how much *more* they would be willing to pay for this programme compared to the lowest ranked programme. A similar question was asked for the programme ranked first.³ Since phrasing the questions in this way reduces the number of possible preference reversals (see Shackley and Donaldson, 2002), the expectation was that the marginal approach would result in greater consistency between the explicit ranking and WTP values. However, the results of the study revealed that there was no statistically significant difference in the proportion of inconsistent respondents between the standard questionnaire (i.e. where partial WTP values were elicited for each programme) and the marginal questionnaire. This finding led the authors to conclude that on the basis of the quantitative evidence, doubt must be cast on the continued use of WTP in this context. However, evidence from the relatively few qualitatively-focused questions pointed to the possibility that the inconsistencies may have been attributable to potential design flaws in the marginal questionnaire causing respondents to misunderstand or misinterpret what was being asked of them. This led the authors to recommend that before abandoning the marginal approach, further research is undertaken which addresses the issues of questionnaire design and includes more qualitatively-focused questions to investigate more thoroughly the reasons underlying respondents' answers. The study reported below represents an attempt to address these issues.

Methods

The questionnaire

The marginal questionnaire used in the Ireland EuroWill survey was adapted to include both programme expansions and reductions. Specifically, the questionnaire contained descriptions of two programme expansions – heart operations and community care services for the elderly – and two programme reductions – hip replacement operations and cancer treatments (the full programme descriptions are in Appendix A). Following some introductory information in which the purpose of the questionnaire and the concept of WTP were explained, respondents were presented with four cards, each of which contained a programme description, and were asked to rank them in terms of their perceived value of each programme. Eight rank orderings were possible, these being 1234, 1134, 1224, 1114, 1233, 1133, 1222 and 1111.

Following the ranking exercises, respondents were asked to state their WTP for each of the four programmes (the actual format of the WTP part of the questionnaire was determined by how the programmes were ranked – there were eight different formats, each corresponding to a particular rank ordering). As with the Ireland survey, the order in which programmes were valued began with the least preferred programme and ended with that which was most preferred. WTP values for the three more highly ranked programmes were elicited using a marginal format, i.e. asking respondents how much more they would be willing to pay relative to a reference value. A payment card with values ranging from £0 to £200 was used to value the least preferred programme, with open-ended questions being used to value subsequent programmes.

² An example of an inconsistency would be someone who ranked programme one ahead of programme two, and both of those ahead of programme three (a 123 ranking) but who stated WTP values of, say, £20, £10 and £20 respectively.

³ This describes the procedure for a distinct 123 ranking. Variations were used for different patterns of ranking.

As well as including proposed programme reductions, the questionnaire also included more qualitative type questions. In addition to asking for reasons why respondents were or were not willing to contribute to one or more of the programmes respondents were also explicitly asked to explain apparent inconsistencies between their WTP values and their ranking of the programmes. At the end of the WTP section, respondents whose implied and explicit ranks were inconsistent with one another were asked to explain the inconsistency. They were also asked which (if either) of the explicit and implied ranks they would prefer a health authority or the government to use to help decide how much of each programme to provide, and to give reasons for their preference. Finally respondents were asked to indicate if (and why) they found WTP questions difficult to answer and to provide a number of socio-demographic details.

Valuing gains and losses from a policy change

When evaluating a proposed policy change within the context of conventional CBA, it is typically the case that some people will gain from the change (i.e. the change will have a positive impact on their utility) while others will lose out (i.e. the change will have a negative effect on their utility). The standard CBA decision rule in these circumstances is the Kaldor-Hicks compensation test which states that if the maximum amount that the gainers from a change would be willing to pay for the change is greater than the minimum amount the losers would be prepared to accept in compensation, then the change should go ahead (Kaldor, 1939; Hicks, 1939).⁴ In considering only proposed expansions to existing programmes, the EuroWill studies avoided the need to consider the possibility that one or more individuals would ‘lose’ as a consequence of a change going ahead since it is extremely unlikely that individuals will experience negative impacts on their utility from expansions to the types of health care programmes presented in the surveys.

One of the main implications of including candidates for reduction is that any reconfiguration of health care services will almost certainly result in there being losers as well as gainers. If respondents regard the proposed programme reductions as having a negative effect on their utility, i.e. they regard the programme as conferring a benefit and any reduction in it will reduce that benefit, then *ceteris paribus*, if the programme is reduced they will be a loser. The reason the issue of losers is important is because of the potential implications it has for how the contingent valuation questions are presented to respondents. In the EuroWill studies, the compensating variation version of WTP was used to elicit values, i.e. WTP for a programme expansion (or gain). If programme reductions (losses) are introduced, then theory states that WTA compensation questions must be used to value those reductions. This presents two problems for the design of the questionnaire – one theoretical and one practical.

The theoretical problem concerns the so-called WTP-WTA disparity, whereby WTA values tend to be considerably larger than WTP values for the same ‘good’, which is contrary to theoretical expectations. For example, in theory, if there are no income effects, one would expect the amount an individual is willing to pay for the introduction of a programme to be equal to the amount of compensation the individual would have to be paid to have the same programme removed. If there are income effects, one would expect WTA to be greater than

⁴ This definition of the Kaldor-Hicks criterion, whilst being that most frequently cited in the literature, strictly refers to the Kaldor version only, where consumer surplus is measured using measures of compensating variation and $\sum CV > 0$. The Hicks version of the criterion uses measures of equivalent variation and is as follows: If the maximum amount that the losers would be willing to pay to prevent the change is less than the minimum that the gainers from a change would be willing to accept in compensation to forgo the change, then the change should go ahead, i.e. $\sum EV > 0$.

WTP. However, in practice the amount by which WTA exceeds WTP tends to be greater than would be predicted by theory. Duborg *et al.* (1994) reported that summaries of the evidence on the so-called WTP-WTA disparity show that mean WTA values are typically between two and five times higher than mean WTP values, with the ratio occasionally reaching double figures. The perceived extent of this problem persuaded the National Oceanic and Atmospheric Administration (NOAA) to recommend that in contingent valuation studies of the environment, WTP measures should be used in preference to WTA (Arrow *et al.*, 1993). Of course, a problem with this recommendation is that it essentially specifying that gains be measured using compensating variation (WTP for the change) while losses be measured using equivalent variation (WTP-to-prevent the change going ahead), which is theoretically incorrect.

Even if the WTP-WTA disparity did not exist, using both WTA and WTP in the same marginal questionnaire would still be problematic. To illustrate this, suppose a respondent was presented with three programmes to consider, one programme reduction and two programme expansions. Suppose further that the respondent ranked the programme reduction last. Using the marginal approach, a WTA compensation amount would first be elicited for the programme reduction. The next programme (a proposed expansion) would then be valued using a marginal question in which the respondent would be asked how much more than their WTA value would s/he be willing to pay for the second ranked programme. It is quite conceivable that the respondent would state s/he is willing to pay less than his/her WTA value, particularly if the difference in their strength of preference between the two programmes is not large.⁵ Even if this pattern of responses did not occur, the use of both WTA and WTP questions adds complexity to the evaluative task. For example, not only would the concept of WTP need to be explained to respondents so too would the concept of WTA compensation. This extra complexity may engender confusion among respondents, causing them to lose sight of the evaluative task and thus affecting the validity of their responses.

A potential solution to the problems surrounding the use of WTA compensation to value losses is to be found in Mitchell and Carson (1989). They outline what they call a new property rights approach, which they claim applies to public goods requiring regular payments to maintain them at a given level of quality. Mitchell and Carson identify two dimensions of public goods which they argue define the appropriate property right and consequently which measure of consumer surplus should be used (WTP or WTA). These are whether the good is individually or collectively held, and whether or not a given level of quality is currently accessible.

Collectively held rights occur when the good is available to all contributors and individuals cannot sell their rights of access. Air and water quality are cited as examples of goods to which consumers have collective, non-transferable property rights. The cost of providing such goods to a given level of quality are normally borne by all consumers through taxes, fees, etc.. If the level of payment is not maintained, the quality of provision will often deteriorate. If a certain level of quality is desired but is currently inaccessible, then in order to attain the desired quality level, higher payments must be made. In a contingent valuation exercise, the relevant measure is the amount of money individuals are willing to pay to ensure the quality improvement goes ahead, i.e. their WTP for a perceived benefit.

⁵ Disregarding the unexplained WTP-WTA disparity, if there are income effects, one would expect WTA values to be greater than WTP values for the same 'good'.

The central tenet of the new property rights approach is that when a proposed reduction in quality from that which is currently accessible is considered, asking respondents to consider receiving compensation is not appropriate. Mitchell and Carson (1989) argue that WTA questions do not make sense in situations where individuals are already paying for a collectively-financed good and where the individual's rights of access to that good cannot be sold. In other words, the WTA format is inconsistent with the non-transferable nature of the property right. In such circumstances, the nature of the property right means that in a contingent valuation exercise, individuals should be asked their WTP to maintain the quality of the provision at its current level, .e. their WTP-to-prevent a perceived loss.

In developing the new property rights approach, Mitchell and Carson (1989) clearly had in mind public goods, in particular environmental goods such as air and water quality. While health care is not a public good in the strict economic sense of being non-rival and non-excludable, it can be argued to possess the same characteristics identified by Mitchell and Carson(1989) as being necessary to allow WTP-to-prevent to be substituted for WTA compensation. Health care which is publicly-financed through taxation is, by definition collectively held. To maintain it at its current level of provision it requires regular payments, and individuals making such payments cannot sell their rights of access to health care. Given these characteristics and the arguments put forward by Mitchell and Carson (1989) in articulating their new property rights approach, a legitimate case can be made against the use of WTA in contingent valuation studies of publicly-financed health care.

The new property rights approach has been tested empirically in the context of a public health intervention, namely the addition of fluoride to drinking water supplies to reduce the incidence of dental caries in children (Shackley and Dixon, 2000). The results of the study not only provided further evidence of a significant disparity between WTP and WTA values, but also illustrated a propensity among respondents to reject the property right implied in the WTA questions. The work by Mitchell and Carson (1989) and Shackley and Dixon (2000) thus provides theoretical and empirical evidence to support the legitimacy of valuing proposed reductions using WTP-to-prevent rather than WTA compensation.

Using the values to inform policy

The following simple example illustrates how the WTP values resulting from the survey could be used to aid health care decision making. To simplify the example, it is assumed that the resources released from either of the two programme reductions are sufficient to fund either of the two programme expansions.

The decision rule follows the principles of the Kaldor-Hicks compensation test in that total WTP for the expansions is compared with total WTP-to-prevent (a proxy for WTA) for the reductions. The decision to expand or reduce one or more of the programmes depends upon the net gains or losses from comparing the programmes.

To illustrate, consider a simple numerical example where the total WTP values for programme expansions E1 and E2 are £1000 and £3000 respectively, and the total WTP-to-prevent values for programme reductions R1 and R2 are £2500 and £2000 respectively. The recommended policy decision would be to expand E2 (highest total WTP of all four programmes) and to reduce R2 (lower WTP-to-prevent value of the two reductions). This would result in a net welfare gain of £1000 (£3000 - £2000). The WTP of the other expansion (E1) is less than the WTP-to-prevent of the other reduction (R1) which means that E1 and R1 should remain at their current levels of provision.

Sample

The survey took place in the spring of 2001 and was conducted by a single interviewer. The area from which the sample was drawn comprised three electoral wards in the city of Sheffield in the United Kingdom. The interviewer randomly selected households in each ward and enquired whether or not a member of the household was willing to be interviewed. In the event of consent being given, the interview took place at that time in the person's home. If the person refused to be interviewed, another house was selected at random within the ward. The intention was to interview roughly equal numbers in each ward.

Data analysis

Tests of difference in average WTP between pairs of the four programmes were performed using t-tests and the Wilcoxon signed ranks test. Tests of differences in the proportions of responses within a particular variable were made using the chi-squared test. The software package used was Stata version 6.0 (StataCorp, 1999).

Results

A total of 100 people were interviewed (33 each from two wards and 34 in the third). Twenty four people declined to be interviewed, giving a response rate of 81 per cent. The sample comprised twice as many females as males, with an age distribution ranging from 21 to 78 years. Thirty-eight per cent of respondents were educated up to at least A level, and 52 per cent were in paid employment or were self-employed.

Zero values and protests to the WTP questions

Six respondents indicated they were unwilling to pay for any of the programmes. When prompted for reasons, three respondents stated that they felt they paid enough taxes already and were thus classified as protesters. Reasons such as this are taken to indicate that respondents are protesting to the concept of WTP and were not willing to engage in the hypothetical exercise (other examples from the EuroWill studies of reasons indicative of protesting were 'users should pay' and 'the health service should be more efficient'). The other three respondents all cited insufficient benefit as the reason for their unwillingness to pay, and were thus classified as true zero responders. The three protest zero responders were dropped from further analysis of the WTP data.

Testing for differences in average WTP

Table 1 shows the mean and median WTP values for the four programmes. The results of tests of differences in average WTP for each of the programmes is shown in Table 2. The cancer programme had the highest mean WTP (£23.77) with the hips programme having the lowest (£15.24).

The t-tests revealed significant differences in mean WTP between the cancer programme and the hearts and hips programmes. When non-parametric tests were used, significant differences in average WTP were found between the cancer programme and all the other programmes, and also between the hearts and hips programmes.

Using total WTP to make policy recommendations

The total WTP values for the programmes were as follows:

Hearts (E1)	£1581.00
Community care (E2)	£1753.50
Hips (R1)	£1478.50
Cancer (R2)	£2306.00

Application of the decision rule outlined above gives the following results. The cancer programme should not be reduced as the total WTP to prevent the reduction is greater than either of the WTP values for the two expansions. The community care programme should be expanded at the expense of the hips programme, and the hearts programme should remain at its current level of provision.

The above has been undertaken on the assumption that the resource savings from the two programme reductions are the same and are equal to the resource costs of the two expansions. If this situation does not apply, then the decision rule needs to be adapted to allow for differences in resource costs and savings. Suppose the hearts and community care programmes both cost £1000, the hips programme reduction frees up £2000 and the cancer reduction frees up £1000. The same recommendation would apply in that the cancer programme should not be reduced and community care programme should be expanded at the expense of the hips programme. However, there would be £1000 left over from the hips programme which could, in principle, be used to fund another expansion. In this example, £1000 is enough to fund the expansion in the hearts programme, and therefore that is what would be recommended. If, however, the hearts programme cost more than £1000, the recommendation would be for it to remain at its current level of provision.

Comparing explicit ranking with the ranking implied from WTP responses

The eight possible rank orderings (and the numbers of respondents associated with each rank) were as follows: 1234 (41); 1134 (8); 1224 (10); 1114 (3); 1233 (7); 1133 (6); 1222 (8); and 1111 (4).

In comparing consistency between explicit and implied ranks, three categories of consistency are used – wholly consistent, partially consistent and inconsistent. Wholly consistent respondents are those for whom the explicit ranking of the programme exactly matches the ranking implied from their WTP values. Partially inconsistent respondents are those for whom the explicit ranking does not exactly match their implied ranking but who cannot be defined as inconsistent, e.g. a 1234 rank with corresponding WTP values of £50, £40, £20, £20 could be explained by the respondent's differential strength of preference between the third and fourth ranked programmes being sufficiently weak as to not be reflected in WTP values, i.e. WTP may not be sensitive enough to reflect differences in weak strength of preference. Inconsistent responses are those for which the explicit and implied ranks are clearly at odds with one another, e.g. a 1234 rank with corresponding WTP values of £40, £50, £10 and £20.

Of the 97 respondents for whom the comparison of ranks was made, only seven (7.2%) were inconsistent. These comprised four respondents with a 1111 rank, two respondents with a 1233 rank and one respondent with a 1134 rank.

Two of the 1111 rankers gave equal values for the first three programmes but a higher value for the last (cancer). When asked to explain the inconsistency, one respondent mentioned cancer being a big killer while the other indicated they had friends who had suffered from cancer. A third 1111 ranker gave a lower WTP value for his/her least preferred programme, with all the other programmes being valued equally. His/her reasons for the inconsistency focused on the benefits of the three higher ranked programmes. The last 1111 ranker valued two programmes at £10 and two at £20. In explaining the apparent inconsistency, this person indicated that s/he had focused on his/her perception of the numbers of people who would be treated in the programmes and his/her estimate of the cost of the programmes.

Both of the inconsistent 1233 rankers were willing to pay increasing amounts for each successive programme they were asked to value, i.e. they were willing to pay more for one of their third ranked programmes than the other. When asked to explain the inconsistency one mentioned personal experience with a programme while the other indicated that closer reflection had led them to amend his/her values.

The inconsistent 1134 ranker stated a WTP value of £1 for each programme. When asked about the inconsistency, the respondent stated that s/he paid enough taxes already. This person can be referred to as a non-zero or partial protester. The individual was willing to engage in the exercise by stating a positive WTP value, but also registered a protest at the amount of taxes they pay.

Of the 90 consistent responders, 79 (81.4% overall) were wholly consistent, with the remaining 11 being partially consistent. This latter group comprised 9 respondents with a 1234 rank, one respondent with a 1134 rank and one with 1233 rank.

Eight of the 1234 rankers gave equal WTP values for two of the four programmes, while the ninth stated an equal WTP for three programmes. The question asking respondents to explain the apparent inconsistencies elicited such responses as “Not prepared to increase contribution compared to cancer”, “Elderly services need improving – consider it as important as using cancer services”. “Value it the same”, “Don’t feel I would want to contribute more” and “Can only pay so much for each”.

The partially consistent 1134 ranker explained his/her unwillingness to pay more for the third ranked programme (community care) compared to the fourth ranked programme (cancer) by saying that people should go to social services for their community care needs.

The partially inconsistent 1233 ranker explained his/her unwillingness to pay more for his/her top ranked programme (hips) than the second ranked programme (cancer) by saying that s/he would want to pay the same for hips as for cancer.

The preferred rank for policy and reasons why

The seven inconsistent and 11 partially consistent respondents were asked to indicate which (if either) of their explicit and implied ranks they would prefer to be used for health care policy purposes and to state the reasons for their preference. Ten respondents indicated they would prefer their WTP values to be used in preference to their explicit ranking, compared with four respondents who had the opposite preference (reasons for these preferences are described below). A chi-squared test of difference in the proportions of respondents preferring one rank or the other was not statistically significant ($\chi^2 = 2.571, p=0.109$). Two respondents felt that neither rank should be used (believing that the public’s views are not

important and that the public cannot be trusted to make decisions such as these), while one wanted both ranks to be used (although s/he did not explain how the inconsistency should be dealt with!). One respondent failed to answer the question.

The four respondents who had a preference for the explicit rank cited the following reasons: “Not sure what is ‘best’ value”; “Because they are the views of people”; “Own inability to choose in an ethical way by attaching values to programmes – don’t want to have to choose personally by attaching values”; and “The NHS would not reallocate vast sums of money”.

Five of the 10 respondents who stated they would prefer WTP values to be used gave reasons for their preference. The reasons were: “Gut feeling”; “WTP values better to make decision from”; “WTP values based more on what something is worth from my perspective – only after going through exercise did I start to think about how much an operation is worth”; “That was what I felt was the right value for the programmes”; and “That’s how I valued them”.

Discussion

The primary aim of this study was to investigate further whether the marginal approach has the potential to improve consistency between WTP values and the explicit ranking of the programmes, and to investigate in more depth the feasibility of asking respondents to value not only proposed programme expansion, but also proposed reductions to existing programmes.

Taking the secondary aim first, the study has demonstrated that it is feasible to elicit values for programme reductions as well as expansions using the marginal approach. This is evidenced by a number of factors. The study had a relatively high response rate (81 per cent) and a very low protest rate (three percent compared with 17.5 per cent in the Ireland EuroWill survey). Regarding reported difficulty in answering the WTP questions, the proportions of responders experiencing difficulty in this survey and the marginal component of the Ireland survey are similar (40 per cent and 35.5 per cent respectively). Perhaps of most significance in terms of demonstrating feasibility is the fact that only 7.2 per cent of responders in this survey gave WTP values which were inconsistent with their explicit ranking of the programmes. The corresponding figure for the marginal questionnaire in the Ireland survey was 20.6 per cent. Possible reasons for this difference in inconsistency rates are discussed later on in this section.

With respect to using the values to inform policy decision making, a simple method has been proposed which uses the total WTP values for the programmes. Applying this method to the data from this survey led to the recommendation that the hips programme be reduced, with the freed resources being used to fund an increase in the provision of community care services. The proposed reduction to the cancer programme should not go ahead, which means that the hearts programme should not be expanded but remain at its current level of provision. The potential for WTP values derived using the marginal approach to inform the type of resource allocation decisions faced by policy makers in a publicly-funded health care system, such as the UK NHS, has thus been demonstrated. It should be noted, however, that because the decision rule involves the aggregation of WTP values, distributional problems may arise.

Perhaps the most oft cited criticism of the use of the Kaldor-Hicks criterion in CBA is that it ignores the possible importance of distributional issues. Since WTP is a function of ability to pay, the Kaldor-Hicks test can lead to decisions being made which favour the rich at the expense of the poor, since the former have a greater ability to pay to support any given

strength of preference. To illustrate, consider the following example in which there are two individuals who are faced with the prospect of a change from which they both stand to gain by the same amount in terms of an increase in their utility. If, as the Kaldor-Hicks criterion asserts, this utility change can be approximated by a monetary value, then, *ceteris paribus*, the WTP values of the two individuals will be the same. However, if one individual is richer than the other, then assuming diminishing marginal utility of income, in order to signal the expected gain in their utility, the richer individual will state a higher WTP than the poorer individual since the marginal utility from a pound gained or lost from a richer person is less than that for the poor person. In the absence of any consideration of distributional issues, this higher WTP value will be taken to mean that the rich individual values the change more highly than the poor individual, which is not the case. As first pointed out by Little (1950), differences in the distribution of income mean that simple aggregation of WTP and WTA values can be problematic.

Of course, it is not universally accepted that there is a problem when aggregating WTP values. For example, Pauly (1995) adopts the following line of reasoning to suggest that simple aggregation of WTP across households is sufficient to arrive at a social valuation:

“... if a dollar’s worth of benefits to poor people were worth more to society than a dollar’s worth of benefits to rich people, it follows that society should be redistributing income from rich people to poor people. If we observe, however, that society, which ever decision it makes, does not seem disposed to make further transfers from rich to poor, then we are not justified in asserting that the same society would value health benefits of a given money value more if they go to poor people than to rich people.

To be sure, an analyst’s personal preference might involve more redistribution to the poor. But it is not legitimate to doctor the books in cost-benefit analysis in order to bring about the subterfuge what would not be tolerated in the standard political process.” P.118

If one does not accept Pauly’s view, then a possible solution to this perceived aggregation problem is to assign distributional weights to the WTP and WTA amounts to reflect differences in income. As Brekke (1997) notes:

“... the choice of money as numeraire may be interpreted as a claim that an additional dollar corresponds to the same amount of cardinal utility independent of wealth, income and characteristics of the person who receives the dollar. If we cannot in general make this assumption, then it follows that individual net benefit has to be weighted.” P.118

While it would be a simple matter to incorporate distributional weights into the above decision rule, the derivation of the weights themselves is likely to prove more difficult. Donaldson (1999) has explored possible methods of weighting WTP and WTA to reflect differences in the income distribution and has suggested that a possible solution to the potential difficulties of deriving ‘accurate’ weights is to test the sensitivity of the results to different assumptions about the weights.

Regarding the study’s primary aim of investigating in more depth the reasons why respondents answer the WTP questions in an apparently inconsistent way, it is useful to return to the results of the Ireland EuroWill survey (Shackley and Donaldson, 2002). Here, a number of speculative explanations of why respondents might answer the WTP questions in an inconsistent manner were put forward. One suggestion was that partially consistent responses could be due to respondents having a relatively weak strength of preference between two or more programmes, thus leading them to assign the same WTP value to each.

Responses to the question asking why respondents were not willing to pay more for a higher ranked programme provide some evidence to support this. Responses included: “Value it the same”; “Don’t feel I would want to contribute more”; and “Would want to pay the same as for cancer”. Further evidence comes from the answers to the question asking respondents to explain the apparent inconsistency between their explicit and implied ranks: “Value cancer the same as hips” and “I ranked them differently but thought they were both worth the same – couldn’t split between them”.

Another suggestion put forward in relation to the Ireland survey was that the wording of the questionnaire may have led respondents to believe they were being asked to pay for all of the programmes, thus leading them to come up against a perceived budget constraint when asked to value their highest ranked programme. In the Ireland survey, the marginal WTP questions were phrased as follows:

“How much more would you be willing to contribute each year to expand the programme compared to the programme?”

It was speculated that the term “compared to” may have led respondents to believe they were being asked to pay for all three programmes from their budget rather than any one programme. The example used to illustrate this was of a respondent who has a nominal budget of £50 which they are willing to allocate to health care. If after valuing each programme they are effectively given this budget back (which is what was intended in the interview), then they have the potential to value each programme up to a maximum of £50. However, if they are under the impression that after valuing the first programme, their budget for the subsequent programmes has been reduced by the amount they said they were willing to pay for the first programme, then there is the potential for their WTP values for the second and /or third programmes to be affected. Specifically these values may be lower than would otherwise have been the case.

To counter this possible effect, the term “compared to” was removed from the WTP questions in the current study. In addition, it was made explicit to respondents for each programme evaluation that their ‘budget’ had not been diminished by any WTP values they may have stated for previous programmes. Specifically, the wording was as follows (the example relates to a programme ranked third in a 1234 ranking):

“I would now like you to consider the programme you ranked third – the programme. In considering this programme, I would like you to imagine that you have been given back the amount you said you were willing to pay for the previous programme. So forget about the previous programme and assume we are back to ‘square one’. We’re only concerned with the programme in this section.”

By contrast, the wording used in the Ireland survey was much less specific, merely asking respondents to assume the other two programmes are not available.

A typical pattern of WTP responses indicative of respondents believing they were being asked to pay for all four programmes would involve the individual not being willing to pay any more for the highest ranked programme compared with their next preferred programme. Only three respondents in this survey exhibited such a pattern, and none of the reasons given to explain their values suggested they felt they were being asked to pay for all four programmes. Rather, the reasons were more indicative of them having a relatively weak strength of

preference between the two programmes of concern, e.g. “Would want to pay the same as for cancer”. This contrasts with the Ireland survey where a number of respondents gave reasons suggestive of reaching a budget constraint, e.g. “I haven’t enough money to be paying for all these things “ and “How can I Pay any more when I don’t have any money?”. The absence of such responses in the current survey suggests that the amendments made to the questionnaire to prevent respondents believing they were being asked to pay for all the programmes were successful.

The purchase of moral satisfaction or warm glow was suggested as another possible reason to explain some of the inconsistencies in the Ireland survey (Kahneman and Knetsch, 1992). Moral satisfaction is purchased by stating a positive WTP amount for what the respondent regards as a good cause. It accrues from the signalling of approval through a positive contribution, with the size of the contribution being of secondary importance. It was suggested that this phenomenon might explain why respondents discriminate between the programmes in the ranking exercise but state the same positive WTP value for each. Only one individual in the current survey followed this response pattern, stating a WTP of £1 for each programme having ranked them 1234. However, the reasons stated for his/her responses pointed to him/her protesting against being asked to pay more taxes. It is interesting to note, however, that unlike a typical protester to a WTP survey, this individual stated a positive (albeit nominal) WTP amount for the programmes, the motivation for which may have been a wish to signal his/her approval of a perceived good cause.

A more fundamental reason why some respondents give inconsistent responses is that they perceived the ranking exercise and the valuation of the programmes through WTP to be different from one another. It was speculated in the Ireland survey that different wording in the ranking and WTP questions might have provoked this interpretation among respondents. In that survey, the ranking exercise asked individuals to indicate how important they thought the programmes were (potentially causing respondents to think from the perspective of society) while the WTP questions focused on value (perhaps implying a more individualistic notion of value). In the current survey, the wording was amended with the intention of conveying the notion of individual value in both contexts. Despite this, however, there is still a suggestion from this survey that some respondents did regard the ranking and WTP exercise as asking different things of them. This is particularly evident in the replies to the question asking respondents to explain why they would prefer one rank or the other to be used for policy purposes: “WTP values are better to make decisions from“; “WTP values are based more on what something is worth from my perspective...”; “That was what I felt was the right value for the programmes”; “That’s how I valued them”; and “Just felt different when I had to attach a value”.

If at least some individuals regard the ranking and WTP exercises as being different, then a decision has to be made regarding which should be used for policy purposes in this context. While a definitive answer to this question is not possible from a small scale study such as this, what data there are seem to suggest that WTP may be the preferred option. Not only did 10 of the 18 inconsistent or partially consistent respondents indicate that they would prefer their WTP values to be used for policy making (compared with 4 who preferred the explicit ranking), but there was also some evidence to suggest that the WTP questions prompted respondents to think more carefully about their preferences. For example one respondent stated “Felt more sympathy for cancer patients once I’d thought about it”, while another claimed “...Only after going through the exercise did I start to think about how much an

operation is worth". This issue is clearly an important one and is worthy of further investigation.

In conclusion, at one level the results of this study can be seen as supporting the notion that a marginal approach to asking WTP questions in this context leads to improved consistency of WTP values with contingent ranking of programmes. At another level, however, the results also provide evidence to suggest that some respondents regard the ranking and WTP exercises as being clearly distinct from one another in that they are asking different things of them. If this is the case, then the notion of consistency between the explicit and implied ranks is no longer relevant. If the explicit ranking and WTP exercises are indeed different, then what is relevant is the need to decide which rank is most appropriate for policy purposes. While this study suggests that WTP is to be preferred, its relatively small size means that further research on this issue is clearly warranted.

Appendix A **Programme descriptions**

Heart operations

An expansion in this programme would result in 100 more heart operations being provided each year.

Most of the extra heart patients are men aged between 60 and 70 years. They have chest pain and breathe heavily when strained.

The operation will result in around 75% of the patients being completely free from pain, with the remainder experiencing less pain. Without the operation the patients are expected to live between 8 and 10 years. With the operation they will be expected to live for an extra year on top of this.

There is an operation mortality risk of 1%, which means that 1 in every 100 people who are operated on will die during the operation.

Hip operations

A reduction in this programme would mean around 80 patients who would otherwise have received a hip replacement this year having to wait longer for the operation.

Most patients needing a hip replacement are women aged between 65 and 85 years. These patients experience pain in the hip and have problems walking. They may be able to do light domestic work, but they often find it difficult to go outside. They may also experience pain during the night.

Following an operation most patients will have complete relief from pain and their mobility and physical functioning will be improved. Most patients benefit from the operation for the rest of their lives, but the operation itself is not life extending.

Community care

An expansion in this programme would mean 150 more physically and mentally dependent elderly people would be able to remain in their own homes rather than being admitted to residential care.

The additional community services would be in the areas of home nursing, home help and day care facilities, and would be targeted at high dependency elderly people living at home. The majority of people benefiting from this programme will be women aged 75 years and over.

The expansion of the community care facilities would improve the quality of life of dependent elderly people living at home, and reduce admissions to long-stay care for people currently on the margin between community care and residential care.

Pain treatment for cancer patients

A reduction in this programme would mean around 50 patients who would have received radiotherapy this year being given pain-relieving medicine instead.

Radiotherapy is given to patients with advanced cancer as a means of relieving pain. The average age of the patients needing treatment is 60 years, although there are patients in every age group. Equal number of men and women are eligible for treatment.

Without this treatment, cancer patients would be provided with pain-relieving medicine. Many patients receiving the medicine will not have satisfactory pain relief, while others will experience side effects in the form of tiredness and poor quality of life.

Radiotherapy for these patients is expected to result in good pain relief for around 75% of patients and improved functioning for most patients. The treatment has few side effects. On average the patients will benefit from this treatment in their last year of life. The treatment will not prolong patients' lives.

References

- Arrow, K., Solow, R., Portney, P.R., Lerner, E.E., Radner, R., Schuman, H., 1993. Report of the NOAA Panel of Contingent Valuation. *Federal Register* 58, 4601-4614.
- Brekke, K.A., 1997. The numeraire matters in cost-benefit analysis. *Journal of Public Economics* 64, 117-123.
- Culyer, A.J., 1971. The nature of the commodity health care and its efficient allocation. *Oxford Economic Papers* 23, 189-211.
- Diener, A., O'Brien, B., Gafni, A., 1998. Health care contingent valuation studies: a review and classification of the literature. *Health Economics* 7, 313-326.
- Donaldson, C., 1999. Developing the method of 'willingness to pay' for assessment of community preferences for health care ('EuroWill'). Final Report to the European Commission Biomed 2 Programme.
- Donaldson, C., 1999. Valuing the benefits of publicly-provided health care: does 'ability to pay' preclude the use of 'willingness to pay'? *Social Science and Medicine* 49, 551-563.
- Duborg, W.R., Jones-Lee, M.W., Loomes, G., 1994. Imprecise preferences and the WTP-WTA disparity. *Journal of Risk and Uncertainty* 9, 115-133.
- Hicks, J.R., 1939. The foundations of welfare economics. *Economic Journal* 49, 696-712.
- Kahneman, D.J., and Knetsch J.L., 1992. Valuing public goods: the purchase of moral satisfaction. *Journal of Environmental Economics and Management* 22, 57-70.
- Kaldor, N., 1939. Welfare propositions of economics and interpersonal comparisons of utility. *Economic Journal* 49, 549-551.
- Klose, T., 1999. The contingent valuation method in health care. *Health Policy* 47, 97-123.
- Little, I.M.D., 1950. *A Critique of Welfare Economics*. Oxford University Press, Oxford.
- Mitchell, R., Carson, R., 1989. *Using Surveys to Value Public Goods*. Resources for the future, Washington DC.
- Mitton, C., Donaldson, C., 2001. Twenty-five years of programme budgeting and marginal analysis in the health care sector, 1974-1999. *Journal of Health Services Research and Policy* 6, 239-248.
- Olsen, J.A., 1997. Aiding priority setting in health care: is there a role for the contingent valuation method? *Health Economics* 6, 603-612.
- Olsen, J.A., Smith, R.D., 2001. Theory versus practice: a review of 'willingness to pay' in health and health care. *Health Economics* 10, 39-52.
- Olsen, J.A., Donaldson, C., Pereira, J., 2004a. The insensitivity of 'willingness to pay' to the size of the good: new evidence for health care. *Journal of Economic Psychology* 25, 445-460.

- Olsen, J.A., Kidholm, K., Donaldson, C., Shackley, P., 2004b. Willingness to pay for public health care: a comparison of two approaches. *Health Policy* (forthcoming).
- Pauly, M.V., 1995. Valuing health care benefits in money terms. In Sloan F.A. (Ed) *Valuing Health Care: Costs, Benefits and Effectiveness of Pharmaceuticals and Other Medical Technologies*. Cambridge University Press, Cambridge, pp. 99-124.
- Protière, C., Donaldson, C., Luchini, S., Moatti, J.P., Shackley, P., 2004. The impact of information on non-health attributes on willingness to pay for multiple health care programmes. *Social Science and Medicine* 58, 1257-1269.
- Ryan, M., Scott, D., 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, 237-258.
- Shackley, P., 2003. Using willingness to pay to aid priority setting in publicly-funded health care systems. PhD Thesis, University of Sheffield, 2003.
- Shackley, P., Dixon, S., 2000. Using contingent valuation to elicit public preferences for water fluoridation. *Applied Economics* 32, 777-787.
- Shackley, P., Donaldson, C., 2002. Should we use willingness to pay to elicit community preferences for health care? New evidence from using a 'marginal' approach. *Journal of Health Economics* 21, 971-991.
- StataCorp, 1999. *Stata Statistical Software: Release 6.0*. Stata Corporation, College Station, TX.
- Stewart, J., O'Shea, E., Donaldson, C., Shackley, P., 2002. Do ordering effects matter in willingness to pay studies of health care? *Journal of Health Economics* 21, 585-599.
- Weisbrod, B.A., 1964. Collective-consumption services of individual-consumption goods. *Quarterly Journal of Economics* 78, 471-477.

Table 1 Mean and median WTP for the programmes

Programme	Mean WTP	Median WTP	n
Hearts	£16.30	£6.00	97
Community care	£18.08	£7.00	97
Hips	£15.24	£5.00	97
Cancer	£23.77	£10.00	97

Table 2 Tests of differences in average WTP using t-test and Wilcoxon Signed Ranks test

Programme comparisons	Significance tests		n
	t-test	Wilcoxon	
Hearts & community care	p=0.465	p=0.594	97
Hearts & hips	p=0.650	p=0.044	97
Hearts & cancer	p=0.003	p=0.000	97
Community care & hips	p=0.211	p=0.126	97
Community care and cancer	p=0.019	p=0.000	97
Hips and cancer	P=0.000	p=0.000	97