

Eliciting community preferences for health care using willingness to pay:
is an 'incremental' approach the way forward?

by

Phil Shackley¹ & Cam Donaldson²

¹ Sheffield Health Economics Group, School of Health and Related Research,
University of Sheffield, UK

² Departments of Economics and Community Health Sciences, University of Calgary,
Canada

Paper presented to the UK Health Economists' Study Group, Keble College,
University of Oxford, 3-5 January, 2001

Abstract

In this paper we present the results from a contingent valuation survey in which WTP values are elicited from members of the general population for three disparate health care programmes with a view to using the values to aid priority setting. The study forms part of the EuroWill project. Previous applications of WTP in this context have revealed a substantial discrepancy between the ranking of the programmes implied by respondents' WTP values and their explicit stated ranking of the programmes. To that end, the focus of our study is to investigate whether an alternative way of presenting WTP questions to respondents (an 'incremental' approach) results in more discriminating values than the 'standard' approach. The results of the comparison provide no clear evidence to recommend the adoption of one approach or the other. Possible reasons why the 'incremental' approach failed to produce a higher proportion of WTP values which were consistent with the explicit ranking of the programmes are explored and recommendations made for further research.

1. Introduction

The use of willingness to pay (WTP) to aid public policy decisions is becoming increasingly popular. The technique has become a well established tool for the evaluation of policies relating to safety and the environment (Jones-Lee, 1989; Arrow *et al.*, 1993), and its application to health care evaluation is becoming more widespread (Diener *et al.*, 1998). Despite this increasing volume of health care literature, the majority of the applications have tended to focus on eliciting WTP values from patients for alternative options for treating the same condition (Bishai and Lang, 2000; Blumenschein and Johannesson, 1998; Chiu *et al.*, 1999; Davey *et al.*, 1998; Donaldson *et al.*, 1998; Johnson *et al.*, 1998; Lee *et al.*, 1998; O'Connor *et al.*, 1998; Ryan, 1998; Zethraeus, 1998). The issue of whether WTP can be used to aid health care priority setting at a broader level has been all but ignored. The one exception to date is a study by Olsen and Donaldson (1998) in which members of the public in Northern Norway were asked their WTP for three disparate health care programmes: a helicopter ambulance service; an increase in the provision of heart operations; and an increase in the provision of hip operations. Respondents were presented with descriptions of each programme and told that each programme would cost the same to implement and that only one could be implemented. While the study demonstrated that using WTP in this context was feasible, the authors called for further development of the method before firmer conclusions could be drawn.

Following Olsen and Donaldson (1998), further work has been undertaken to investigate the use of willingness pay in a broader priority setting project under the auspices of a European Commission funded project, which has become known by the short title 'EuroWill' (Donaldson, 1999). The project involved the carrying out of an interview-based WTP survey in six European countries (Denmark, France, Ireland, Norway, Portugal and the United Kingdom). While the general format of each country's survey was broadly similar, insofar as members of the public were asked to set priorities amongst alternative uses of health care resources through the use of WTP, specific issues were also addressed by each of the surveys. This paper reports the results of the survey that was undertaken in Ireland.

The stimuli for the survey are twofold. First, it has been found in other studies that there is a substantial discrepancy between the ranking of the programmes implied by

respondents' WTP values and their explicit stated ranking of the programmes. Second, an alternative approach (termed the 'marginal' approach and described in the next section) has worked well in studies of pairwise choices of treatments made by patients. The principal aim of the Irish survey, therefore, was to investigate whether an alternative method of framing the WTP questions based on the marginal approach resulted in greater consistency between respondents' implied and explicit rankings in this broader priority setting context. The method is termed the 'incremental' approach.

2. Background

2.1 Valuing close substitutes

The origin of the incremental approach lies in previous applications of WTP to the evaluation of 'close substitutes', i.e. alternative options for treating the same condition, where values are elicited from patients. An apparent inability of patients' WTP values to discriminate between treatment options in randomised trials led to the development of a new approach to eliciting WTP values in this context which was termed the 'marginal' approach.

Originally, when used alongside randomised trials, patients were asked about their WTP for the care they received (Donaldson *et al.*, 1995). Those in group A would be asked to provide a value $V(A)$ and those in group B would provide a value $V(B)$. In several studies, it has been shown that $V(A)=V(B)$, and it is speculated that this is because patients in each group compare the value of being treated with receiving no treatment at all (Johannesson and Fagerburg, 1992; Donaldson *et al.*, 1995; Ryan *et al.*, 1997).

Building on reference point theory as exemplified by Kahneman and Tversky (1979) and Schoemaker (1982), Donaldson *et al.* (1997a) used a "WTP for each" approach in which respondents provided values for each option, $V(A)$ and $V(B)$. However, this approach suffered from problems of orderings from WTP values not reflecting respondents' stated preferences, likely due to the fact that respondents tended to estimate the cost of the alternatives evaluated and state that as their WTP. If a preferred option was thought to be less costly, some respondents would give it a lower WTP.

This phenomenon was also reported in a recent study by Ryan and San Miguel (2000) in which respondents to a WTP survey were asked to consider two treatments for menorrhagia – hysterectomy versus conservative treatment. Respondents were instructed to state which option they preferred and a WTP value for each. It was found that 41 per cent of the sample were inconsistent in that their WTP for their less preferred option was greater than the WTP amount for their preferred option. In view of this finding, Ryan and San Miguel (2000) suggested that the ‘marginal’ approach may overcome cost-based responses and hence produce more consistent results.

The ‘marginal’ approach was first articulated by Donaldson *et al.* (1997a) and involves respondents being asked to indicate their preferred treatment option and to state how much they would be willing to pay to have that *rather than* their less preferred option. As well as relating to reference point theory, it could also be argued that this approach fits more closely with the Kaldor-Hicks criterion in that, through elicitation of preferences, "winners" and "losers" from a change in the way care is delivered are identified and their strengths of preference for their preferred option (over and above that which is less preferred) compared. The approach was first applied to child health services, and subsequently to maternity care and choices about management of miscarriage (Donaldson *et al.*, 1997b; Donaldson *et al.*, 1998; Gibb *et al.*, 1998). In these contexts, the approach seems to have worked well. For example, it has been shown in one study that respondents who stated a preference for a less costly option, had a higher WTP for that option on average, indicating that the role of cost considerations in valuations is diminished (Donaldson *et al.*, 1997b). A more detailed discussion of the approach can be found in Shackley and Donaldson (2000).

2.2 *Eliciting WTP values from the public*

The marginal approach has subsequently been applied to water fluoridation (Shackley and Dixon, 2000), where, again, respondents were asked to choose between two alternatives and to place an incremental value on that which was preferred. These respondents were members of the general public, although the options were still relatively narrow (i.e. two different ways of improving oral health).

Prior to this, other studies eliciting values from the general public for more than two broad options (e.g. helicopters, hearts and hips) revealed that, without using the incremental approach, respondents' stated rankings of options did not closely match the ranking implied by their WTP values (Olsen, 1997). A similar result has been shown in two EuroWill surveys (Olsen *et al.* 2000; Donaldson *et al.*, 2000).

Combining the problems revealed in the Olsen study with the promise shown in the application of the marginal approach to choices over pairs of goods, it was hypothesised that by making the relationship between the respondents' stated ranking of the programmes and their WTP values more explicit, the incremental approach would result in greater consistency between the two.

3. Methods

3.1 The 'standard' and 'incremental' approaches

Two questionnaires were used in the Irish survey, a 'standard' EuroWill questionnaire and an 'incremental' questionnaire. Both questionnaires had the same introductory information in which the purpose of the questionnaire was explained and the respondent was introduced to the three programmes to be valued – cancer treatments, heart operations and community care services for the elderly. Respondents were then asked about their self-perceived risk of ever needing each of the three programmes and whether anyone in their close family had ever had first hand experience of cancer, heart disease or community care services. Next, respondents were presented with written descriptions of each of the three programmes and asked to rank them in terms of how important they thought each programme was. All programmes were described in terms of expansions to existing services (the full programme descriptions are in the Appendix). In the next section, respondents were asked about their WTP for each of the programmes. The questionnaire finished with some demographic questions.

It was with regard to the WTP section that the two questionnaires differed. In the standard version respondents were first asked about the cancer programme and told to disregard the other two programmes. They were then asked if they would be willing to contribute anything in extra taxation for the proposed expansion in cancer treatments (to allow for the fact that some respondents would not be tax payers or

may have had an aversion to paying taxes, they were also asked if they would be willing to pay by means of a voluntary donation). Those who were not willing to pay were asked to state why, while those who were willing to pay were asked how much. The WTP values were elicited with the help of a payment card which listed a number of values from £0 to £200, from which respondents were asked to indicate the value that corresponded to their maximum WTP (respondents whose WTP was over £200 had the option of writing their value in a space provided). Finally, respondents were asked to indicate from a list of possible reasons why they were willing to pay for the programme. This exercise was then repeated for the hearts programme and the community care programme.

In the incremental questionnaire, following the ranking exercise, the lowest ranked programme was selected for the first WTP valuation (in the event that two or more programmes were ranked jointly, the order in which programmes were selected was cancer, hearts, community care). Assuming for now a discrete '1', '2', '3' ranking, the WTP value for programme ranked third was elicited in exactly the same way as in the standard questionnaire. The next programme considered was that ranked second. Respondents who indicated a WTP for this programme were presented with the following question:

You ranked the..... programme ahead of the programme, and you were willing to pay £..... to expand the programme. How much MORE would your household be willing to contribute each year to expand the programme compared to the programme?

(Note that the interviewer inserted the appropriate programmes and WTP value in the spaces provided).

The same payment card as that in the standard questionnaire was used. To check that respondents had understood the incremental nature of the exercise, the interviewer presented them with the total WTP amount their stated incremental amount implied they were willing to pay for the programme. Any respondents who disagreed with the total WTP figure were asked to state what their total WTP was. If respondents were not willing to pay any more than they were for the first programme,

they were asked to state why. As with the standard questionnaire, reasons why respondents were willing to contribute were also elicited.

The valuation procedure for the first ranked programme was exactly the same as that described for the programme ranked second except that the comparator programme for the incremental WTP amount was the programme that was ranked third (not the second ranked programme). This was always the case for the programme valued last.

Variations of the above valuation procedures were employed for different rankings of the programmes, i.e. two programmes ranked joint second (122), two programmes ranked joint first (113), and all programmes ranked equally (111). For the incremental questions when programmes were ranked jointly with the comparator programme, i.e. rankings of 122 or 111, respondents were asked whether they would be willing to pay the same, more or less for the relevant programme expansion. If respondents answered the same or less they were asked to explain why.

As pointed out at the end of the previous section, it was hypothesised that by making the relationship between the respondents' stated ranking of the programmes and their WTP values more explicit in the incremental version, a greater consistency between the two would result. It should be noted that the incremental design of the questionnaire gives WTP its 'best shot' at producing values which are consistent with the explicit ranking of the programmes.

3.2 *Sample*

The survey took place during the summer of 1999 and was conducted by the Economic and Social Research Institute (ESRI), Ireland's primary economic and social research agency. The study area was the Western Health Board region of Ireland which has an approximate population of 350,000 and contains the counties of Galway, Mayo and Roscommon. The sample was generated by means of two-stage cluster sampling using the electoral register as a population frame. A preliminary sample was initially selected and then adjustments made to ensure that it was representative of the population from which it was drawn.

3.3 *Data analysis*

Tests of differences between the two samples were made using chi-squared tests. Within and between sample tests of differences in WTP for each of the programmes were performed using t-tests and non-parametric tests (the Mann-Whitney test for between sample comparisons and the Wilcoxon Signed Ranks test for within sample comparisons). The relationship between respondents' explicit ranking of the programmes and the implied ranking from their WTP values was investigated and chi-squared tests of differences performed where appropriate. Finally, a series of ordinary least squares (OLS) regression analyses were performed in which WTP and $\ln(\text{WTP})^1$ were regressed against a number of explanatory variables. For each regression, appropriate tests of misspecification were performed.

4. **Results**

A total of 234 people were interviewed of whom 113 received the standard questionnaire and 121 received the incremental version. The numbers of people who refused to be interviewed were 16 for the standard questionnaire and 15 for the incremental questionnaire. Thus the 'response rates' to the two questionnaires were 88% and 90% respectively.

There were no significant differences between the two groups of respondents in terms of demographic variables (age, sex, education, gross annual income, marital status, employment status, smoking habits and drinking habits). There were also no differences between the questionnaires in terms of respondents' experience of cancer, heart disease or community care services, nor were there any differences in their self-perceived risk of requiring any of the three programmes. A difference was found with respect to the respondents' reported difficulty in answering the WTP questions, with a significantly higher proportion of respondents to the incremental questionnaire finding the WTP questions difficult ($\chi^2 = 8.452$, $p=0.004$).

One respondent to the incremental questionnaire who ranked all three programmes equally gave a negative WTP value for the third programme they valued (community care). This came about from them stating they were willing to pay £100 less for this

programme compared to their WTP for the cancer programme (£60). When they were confronted with the fact that their total WTP for community care was minus £40 they still persisted with the negative value. In view of this, it was decided to drop this respondent from further analysis on the grounds that they had not understood what was being asked of them.

4.1 Zero values and protests to the WTP questions

The numbers and proportions of true zero, protest zero and positive WTP responses to each of the three programmes by type of questionnaire are presented in Table 1. A true zero response is defined as one where a respondent answered 'The programme is of no value to my household' or 'I can't afford it' to the question of why they were not willing to pay anything for the programme expansion. A protest zero response is defined as one where a respondent cites such reasons for not being willing to pay as 'Other public sector budgets should be cut', 'The users should pay', or 'The health service should be more efficient'. Such reasons are taken to indicate that the respondent is protesting to the concept of WTP and is refusing to engage in the hypothetical exercise. Chi-squared tests for differences in the proportions of true zeros, protest zero and positive values within each programme revealed no statistically significant differences between the questionnaires.

4.2 Testing for differences in average WTP

The results of the within sample tests of differences in average WTP for each of the programmes are shown in Tables 2 and 3. The tests were performed for three different samples of data: (i) all zero values included; (ii) true zero values only included; and (iii) no zero values included, i.e. positive values only. In all cases, the cancer programme had the highest mean WTP value while the community care programme had the lowest.

For all three samples of data, in the standard questionnaire, t-tests revealed significant differences between mean WTP for the cancer programme and the other two programmes, but not between the hearts and community care programmes.

¹ The actual dependent variable was $\ln(WTP+1)$. The addition of £1 to the WTP values was necessary because the natural logarithm of zero is unspecified.

When non-parametric tests were used, significant differences in average WTP were found between all pairs of programmes.

In the incremental questionnaire, again for all three samples of data, t-tests revealed significant differences in mean WTP for the community care programme compared to the other two programmes. As was the case with the standard questionnaire, when non-parametric tests were performed, significant differences were found in average WTP between all pairs of programmes.

The results of the between sample tests of differences in average WTP for each of three programmes are shown in Table 4. The tests were performed for the same three samples of data as were used for the within sample tests of differences in average WTP. Regardless of which sample was used, none of the t-tests produced a significant result for any of the programmes. Similarly, the non-parametric tests revealed no significant difference between the standard and incremental questionnaires in terms of the average WTP for the cancer programme. However, average WTP values for hearts and community care were both significantly higher in the incremental questionnaire when the samples included all zero values and when the samples included true zeros only. No significant differences were found when positive WTP values only were considered.

4.3 Comparing explicit ranking with the ranking implied from WTP responses

Table 5 shows the distribution of responses to the explicit ranking questions by type of questionnaire. The pattern of rankings between the questionnaire is the same for the second and third ranks, but differs with respect to the programmes ranked first. In the standard questionnaire, the cancer programme was ranked first most frequently, followed by the community care programme then the hearts programme. While the cancer programme was also ranked first most frequently in the incremental questionnaire, the next most frequent first ranked programme was the hearts programme. Chi-squared tests of differences in the distribution of respondents' answers to the ranking questions revealed no significant differences between the questionnaires.

As indicated earlier, four different rank orderings of the programmes were possible: one programme is preferred to another which is in turn preferred to the third (123); two programmes are ranked jointly and are preferred to the third (113); one programme is preferred to the other two, which are both ranked jointly (122); and all programmes are ranked equally (111). Table 6 shows the distribution of these different ranking possibilities between the questionnaires for all respondents. A significant difference in the distribution of ranks between the questionnaires was found ($\chi^2 = 9.916$, $p=0.019$). A series of further 2 x 2 chi-squared tests revealed that the incremental questionnaire had a higher proportion of 113 type ranks than the standard questionnaire ($\chi^2 = 8.897$, $p=0.003$). A similar result was found when protesting individuals were excluded, i.e. a significant difference in the distribution of ranks was found ($\chi^2 = 8.114$, $p=0.044$) and the incremental questionnaire had a higher proportion of 113 ranks ($\chi^2 = 7.668$, $p=0.006$).

In order to investigate the consistency between respondents' explicit ranking of the programmes and the ranking implied by their WTP values, four different groupings of consistency were defined. Wholly consistent respondents were those for whom the explicit ranking of the programmes was identical to their implied WTP ranking, e.g. a rank of 123 might have corresponding WTP values of £30, £20 and £5, or a rank of 111 might have corresponding WTP values of £10, £10, and £10. Inconsistent respondents were defined as those whose WTP values were clearly at odds with their explicit ranking of the programmes, e.g. a rank of 123 might have corresponding WTP values of £60, £100 and £50, or a rank of 111 might have corresponding WTP values of £20, £10 and £0. Partially consistent respondents were those for whom the explicit ranking of the programmes did not exactly match their implied WTP ranking, but who could not be defined as inconsistent, e.g. a rank of 123 might have corresponding WTP values of £30, £20 and £20, or a rank of 122 might have corresponding WTP values of £15, £15 and £15. A possible explanation for such patterns of responses might be that a respondent's strength of preference might be sufficiently weak as to not be reflected in their WTP values, i.e. WTP is not a sensitive enough measure to reflect differences in weak strength of preference. Two groups of partially consistent responses were identified depending on whether or not all three WTP values were equal.

Table 7 shows the distribution of wholly consistent, partially consistent and inconsistent responses between the standard and incremental questionnaires for each of the four possible explicit ranking combinations (protesting individuals excluded). The proportion of wholly consistent responses is higher in the incremental questionnaire for all but one of the explicit rankings, the exception being 111 ranking where the standard questionnaire has a higher proportion of consistent responses. The proportion of inconsistent responses is also higher in the incremental questionnaire for three of the four possible rankings, the exception here being the 122 ranking where neither questionnaire registered an inconsistent response. However, when tests for significant differences in the distribution of consistent responses between the two questionnaires across the four rankings were performed, none was found. This finding was replicated when all combinations of rankings were grouped together, i.e. there is a higher proportion of both consistent and inconsistent responses in the incremental questionnaire, but the differences are not significant.

4.4 Regression analyses

The independent variables included in the regression analyses are listed in Table 8. Regression equations were initially estimated with WTP as the dependent variable. However, the three resultant regression models were found to suffer from misspecification problems and heteroscedasticity. Subsequent estimation in which the dependent variables were log-transformed WTP produced fewer specification problems (Table 9). Indeed, only one of the six estimated models was problematic. This was the model for the cancer programme from the incremental questionnaire, where evidence of heteroscedasticity was found.

Considering the signs of the coefficients, the effects on WTP of SEX, EDUC, DIFFIC and INTEREST are the same for all six models, i.e. women are willing to pay more than men, better educated respondents are willing to pay more, those who found the WTP questions difficult are willing to pay more, as are those who were more interested in the exercise. INTEREST is highly significant in the models estimated from the standard questionnaire, but insignificant in the incremental models.

In all but one of the models, the coefficients on AGE indicate that older respondents are willing to pay less for the programmes. The exception is the standard questionnaire model for the community care programme where the reverse is true. It is interesting to note that AGE is a significant predictor of WTP only in the models derived from the incremental questionnaire.

The coefficients on HEALTH indicate that in five out of six cases, respondents whose self-assessed health is good or very good are willing to pay more than their less healthy counterparts. The exception here is again the standard questionnaire model for the community care programme. The only case where HEALTH is a significant predictor of WTP is in the community care model for the incremental questionnaire.

Respondents who have had first hand experience of cancer and community care services are willing to pay more for these programmes than those who have not experienced them. The reverse is true for the hearts programme. EXP is a significant predictor of WTP only in the community care model for the incremental questionnaire.

The effects of self-assessed risk of ever needing the health care programmes on respondents' WTP vary across programmes and questionnaires. As far as significant effects are concerned, respondents whose assessment of their risk is above average are willing to pay more for the hearts programme and less for the community care programme (standard questionnaire models only).

If one accepts the view that WTP is a function of ability to pay, then one would expect the coefficients on the five income variables to be positive and to show an increasing trend from INC_A to INC_E. This is the case for the incremental models for hearts and community care and the standard model for cancer. The coefficients in the standard heart model follow this pattern for INC_A to INC_D, but INC_E has a negative coefficient indicating that respondents in the highest income bracket are willing to pay less than those in the lowest bracket. In the standard hearts model, while the coefficient on INC_E is positive, it is smaller than those for INC_A to INC_D which is contrary to expectations. In the incremental cancer programme, three of the five income variables have negative coefficients, supporting the finding that this

model is misspecified. It is interesting to note that income is a significant predictor of WTP in both the hearts and community care programmes for the incremental questionnaire, but only in the community care programme for the standard questionnaire.

5. Discussion

The aim of this paper has been to compare two alternative methods of eliciting WTP values from the general public – a standard approach and an incremental approach. An interesting result from the study is the finding that mean WTP for all three programmes was higher in the incremental version than in the standard version. One interpretation of this result is that the incremental approach is better at extracting consumer surplus from respondents and therefore a better measure of maximum WTP. However, while the differences in average WTP were statistically significant for the hearts and community care programmes, the difference was not significant for the cancer programme. A possible explanation for this may be the order in which the programmes were valued in the incremental questionnaire. Only the programmes valued second and third had WTP amounts elicited using an incremental question – the WTP amount for the programme valued first was elicited in the same way as in the standard questionnaire. Excluding protesters, the cancer programme was valued 55 times by an incremental question compared with 72 times for the community care programme and 86 times for the hearts programme. If the higher mean WTP figures in the incremental questionnaire are attributable to the incremental questions, then the fact the cancer programme was valued by an incremental question on the fewest occasions may account for the non-significant difference in average WTP.

One of the principal aims of the study was to investigate whether the incremental version of the questionnaire resulted in greater consistency between the respondents' explicit ranking of the programmes and the ranking implied from their WTP values. Tests of differences in proportions of consistent responses between the two questionnaires revealed no significant differences, thus suggesting that the incremental version is no better than the standard version at providing WTP responses which are more closely related to respondents' explicit ranking of the programmes.

In view of the fact that the incremental questionnaire was designed with the express purpose of producing more consistent responses, it is interesting to speculate as to why it apparently failed to do so. A contributory factor may have been the way the WTP questions were framed in the incremental questionnaire. The wording was as follows, "How much more would you be willing to contribute each year to expand the programme compared to the programme?". It is possible 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. An alternative term which might have conveyed more clearly to respondents that they were not being asked to pay for all three programmes is "rather than" or "instead of", as was used in earlier studies of patients (Donaldson *et al.*, 1997; Donaldson *et al.*, 1998). If respondents were under the impression they were being asked to pay for all three programmes, then this had the potential to affect WTP values and consequently consistency between explicit and implied ranks. Consider the example 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 an/or third programmes to be affected. Specifically, these values may be lower than would otherwise have been the case. To investigate whether this occurred in the survey respondents would ideally have been asked to explain any apparent discrepancies that occurred in their WTP values. Unfortunately, no such question was included in the incremental questionnaire.² An alternative (albeit far from ideal) method is to compare patterns of WTP responses between the two questionnaires. This was done by considering respondents who had a distinct 123 rank ordering of the programmes. In particular, the analysis focused on those whose WTP values discriminated between the third and second ranked programmes, but whose WTP value for the first ranked programme was less than or equal to that for the second

² Respondents were asked for reasons why they were not willing to pay any more than previously stated, but were not asked to explain why they were willing to pay a positive, but smaller amount than previously.

ranked programme. In the standard questionnaire, 10 respondents followed the above pattern of WTP responses. Of these, nine had a WTP value for the first ranked programme that was equal to their WTP for the second ranked programme. This was also the number in the incremental questionnaire. However, seven respondents to the incremental questionnaire stated a WTP value for the first ranked programme which was less than that for the second ranked programme. This is compared to only one such respondent in the standard questionnaire. While this may be regarded as evidence to support the notion that respondents to the incremental questionnaire believed they were being asked to pay for all three programmes, a chi-squared test of difference in proportions between the two questionnaire was not significant ($\chi^2 = 3.291$, $p=0.070$), although the numbers were small.

A budget constraint of a different type may also have affected consistency. Suppose a respondent with a 123 ranking has stated a WTP of £10 for each of the programmes. The apparent inconsistency may be due to the respondent deciding that the most they can afford to contribute to any single programme is £10. If they value the third ranked programme at at least £10, then their budget constraint prevents them from differentiating between the second and first ranked programmes in terms of WTP. They may want to contribute £20 and £30 to the second and first ranked programmes respectively, but they cannot do so because of their budget constraint means they cannot afford to.

Another potential explanation for why the incremental questionnaire failed to provide more consistent responses is the phenomenon known as the purchase of moral satisfaction or warm glow, whereby respondents who regard a programme as a good cause wish to indicate their approval of it through a positive WTP amount (Kahneman and Knetsch, 1992; Schkade and Payne, 1994). Moral satisfaction accrues from being seen to make a contribution, the actual size of the contribution being of secondary importance. This phenomenon has been put forward as an explanation of why respondents in WTP surveys fail to vary their WTP values as the magnitude of benefit associated with a single programme changes, e.g. if the number of people benefiting from the hearts programme doubled, one would expect

a respondent's WTP to increase accordingly. Failure to discriminate between the two magnitudes of benefit is referred to as an embedding effect. Moral satisfaction can be purchased with the same WTP amount regardless of the size of the benefit. While this study is not concerned with the embedding effect specifically, it is possible that a type of embedding effect could have occurred. All three programmes can be regarded as a good cause, i.e. expansions in health care provision. Respondents may have wanted to signal their approval of expanding health care provision generally through being seen to make a positive contribution towards such expansions. Similar to embedding, respondents may have considered that what was important was to make a positive contribution, with the amount of the contribution not varying across the different programmes.

Another contributory factor which has already been alluded to above is the possibility that WTP may not be sensitive enough to reflect differences in weak strength of preference. For example, if a respondent decides that they prefer the cancer programme to the hearts programme to the community care programme, it is relatively straight forward to reflect this preference with a 123 ranking of the programmes. However, if the respondent's preferences between the programmes is relatively weak, then they may feel that they are unable to distinguish between the programmes in terms of WTP. This phenomenon may be exacerbated by the use of a payment card which encourages respondents to select values in increments of £5, £10 or £20 rather than stating values in between, e.g. it is quite conceivable that if the range of values the respondent is willing to pay for each of the three programmes is between £30 to £35 they will select £30 for each programme.

A more fundamental explanation of the inconsistency between the explicit ranking of the programmes and the implied ranking from WTP values is that the ranking and WTP questions are asking two different things of respondents. The ranking exercise asks respondents to rank the programmes in terms of how *important* they think they are, while the WTP exercise asks respondents to consider how much they *value* each one. The use of the word "important" could be construed as indicating a societal setting whereas "value" could be taken to imply a more individualistic context. Clearly, societal and individual preferences need not coincide. For example, an individual may regard community care services as the most important

programme from a societal point of view, but may value the hearts programme more highly from an individual perspective, perhaps because there is a history of heart disease in their family. If the explicit ranking and WTP exercises are different, then the notion of consistency between the two is no longer a relevant means of comparing the two questionnaires. Another, more straightforward explanation, of course, is that, if respondents are asked to rank options, they will do so, whilst their WTP values are less precise.

In light of this, an alternative method of comparing the questionnaires was sought. It was decided to investigate whether there were any differences in the rankings implied by WTP values and the extent to which respondents discriminate between the programmes in terms of WTP values. With respect to the rankings, the incremental questionnaire had higher proportion of respondents with implied ranks of 123, 113 and 122 and a lower proportion of 111 ranks. Fifty-three per cent of respondents to the incremental questionnaire discriminated between programmes at least once with their WTP values compared with 40 per cent in the standard questionnaire. It may be tempting to conclude on the basis of this evidence that the incremental questionnaire produced more discriminating WTP responses. However, a chi-squared test of differences in proportions of ranks between the questionnaire was not significant ($\chi^2 = 4.034$, $p=0.258$), nor was a chi-squared test of differences in respondents whose WTP values discriminated between the programmes ($\chi^2 = 3.216$, $p=0.073$).

The final way in which the two questionnaires were compared was with regression analysis. The higher R-squared statistics in the incremental models and the trend and significance of the income variables could be regarded as evidence to support the incremental approach. However, this must be set against the fact that the only model which suffered from specification problems was the incremental cancer model.

Overall, the results of the comparison of the standard and incremental questionnaires provide no clear evidence to recommend the adoption of one approach or the other. In view of the fact that the design of the incremental

questionnaire gave WTP its 'best shot' at producing values which were consistent with the explicit ranking of the programmes, failure to produce a higher proportion of consistent WTP values could be regarded as a worrying result for the technique. That said, the discussion has highlighted possible reasons why the incremental questionnaire failed to produce a higher proportion of WTP values which were consistent with the explicit ranking of the programmes, but much of it has been speculative in nature. The approach does show some potential, and it remains puzzling why it did not work as well in this context relative to how it has appeared to work in earlier studies. In view of this, there would seem to be a clear need for future testing of different forms of the incremental approach, including the use of a more qualitative approach in order to investigate more thoroughly the reasons underlying respondents' answers.

Table 1 Numbers (and percentages) of true zero, protest zero and positive WTP responses to each of the three programmes by questionnaire

	Standard questionnaire			Incremental questionnaire		
	Cancer	Hearts	C Care	Cancer	Hearts	C Care
True zeros	11 (9.7%)	15 (13.3%)	13 (11.5%)	6 (5.0%)	8 (6.7%)	8 (6.7%)
Protest zeros	11 (9.7%)	14 (12.4%)	17 (15.0%)	13 (10.8%)	12 (10.0%)	14 (11.7%)
Positive values	91 (80.5%)	84 (74.3%)	83 (73.5%)	101 (84.2%)	100 (83.3%)	98 (81.7%)

Table 2 Within sample tests of differences in average WTP using t-tests and Wilcoxon Signed Ranks test (standard questionnaire)

Sample	WTP measure	Programme						Within sample comparisons of average WTP					
		Cancer	n	Hearts	n	C Care	n	Cancer & Hearts		Cancer & C Care		Hearts & C Care	
								t-test	Wilcoxon	t-test	Wilcoxon	t-test	Wilcoxon
All zeros included	Mean WTP	£44.42	113	£37.12	113	£31.55	113	p=0.019	p=0.000	p=0.000	p=0.000	p=0.102	p=0.016
	Median WTP	£20.00	113	£10.00	113	£10.00	113						
True zeros only included	Mean WTP	£49.22	102	£42.37	99	£37.14	96	p=0.019	p=0.000	p=0.001	p=0.000	p=0.083	p=0.029
	Median WTP	£20.00	102	£20.00	99	£10.00	96						
No zeros included	Mean WTP	£55.16	91	£49.94	84	£42.95	83	p=0.042	p=0.002	p=0.001	p=0.000	p=0.074	p=0.018
	Median WTP	£25.00	91	£20.00	84	£20.00	83						

Table 3 Within sample tests of differences in average WTP using t-tests and Wilcoxon Signed Ranks test (incremental questionnaire)

Sample	WTP measure	Programme						Within sample comparisons of average WTP					
		Cancer	n	Hearts	n	C Care	n	Cancer & Hearts		Cancer & C Care		Hearts & C Care	
								t-test	Wilcoxon	t-test	Wilcoxon	t-test	Wilcoxon
All zeros included	Mean WTP	£48.40	120	£45.56	120	£37.17	120	p=0.170	p=0.049	p=0.000	p=0.000	p=0.009	p=0.010
	Median WTP	£25.00	120	£20.00	120	£20.00	120						
True zeros only included	Mean WTP	£54.28	107	£50.63	108	£42.08	106	p=0.116	p=0.026	p=0.001	p=0.000	p=0.028	p=0.020
	Median WTP	£30.00	107	£25.00	108	£20.00	106						
No zeros included	Mean WTP	£57.50	101	£54.68	100	£45.52	98	p=0.138	p=0.037	p=0.003	p=0.000	p=0.047	p=0.033
	Median WTP	£40.00	101	£30.00	100	£25.00	98						

Table 4 Between sample tests of differences in average WTP using t-tests and Mann-Whitney U tests

Programme and sample	Standard questionnaire			Incremental questionnaire			Significance tests	
	Mean WTP	Median WTP	n	Mean WTP	Median WTP	n	t-test	Mann-Whitney U
Cancer								
- All zeros included	£44.42	£20.00	113	£48.40	£25.00	120	p=0.604	p=0.132
- True zeros only included	£49.22	£20.00	102	£54.28	£30.00	107	p=0.539	p=0.064
- No zeros included	£55.16	£25.00	91	£57.50	£40.00	101	p=0.788	p=0.170
Hearts								
- All zeros included	£37.12	£10.00	113	£45.56	£20.00	120	p=0.268	p=0.032
- True zeros only included	£42.37	£20.00	99	£50.63	£25.00	108	p=0.320	p=0.037
- No zeros included	£49.94	£20.00	84	£54.68	£30.00	100	p=0.599	p=0.179
Community Care								
- All zeros included	£31.55	£10.00	113	£37.17	£20.00	120	p=0.400	p=0.031
- True zeros only included	£37.14	£10.00	96	£42.08	£20.00	106	p=0.506	p=0.039
- No zeros included	£42.95	£20.00	83	£45.52	£25.00	98	p=0.750	p=0.118

Table 5 Numbers of respondents ranking the programmes first, second and third

Rank	Standard questionnaire			Incremental questionnaire		
	Cancer	Hearts	C Care	Cancer	Hearts	C Care
1 st	90	54	59	93	72	63
2 nd	20	40	12	22	31	12
3 rd	3	19	42	6	18	46
n	113	113	113	121	121	121

Tests of differences in the distribution of respondents' answers to the ranking questions:
 Cancer ($\chi^2 = 0.872$, $p=0.647$); Hearts ($\chi^2 = 3.470$, $p=0.176$); Community Care ($\chi^2 = 0.040$, $p=0.980$)

Table 6 Distribution of different programme rankings between the questionnaires

Ranking	Standard questionnaire		Incremental questionnaire	
	Number	Percentage	Number	Percentage
123	56	49.6%	45	37.2%
113	8	7.1%	25	20.7%
122	8	7.1%	10	8.3%
111	41	36.3%	41	33.9%

Test of differences in the distribution of the rankings: ($\chi^2 = 9.916$, $p=0.019$)

Table 7 Distribution of wholly consistent, partially consistent and inconsistent responses between the standard and incremental questionnaires for each of the four possible explicit ranking combinations

	Explicit ranking 123		Explicit ranking 113		Explicit ranking 122		Explicit ranking 111		All rankings	
	Standard	Increment	Standard	Increment	Standard	Increment	Standard	Increment	Standard	Increment
Wholly consistent	7 (16.3%)	7 (17.5%)	0	8 (38.1%)	2 (40.0%)	3 (42.9%)	29 (80.6%)	26 (76.5%)	38 (42.2%)	44 (43.1%)
Partially consistent - WTP values not all equal	16 (37.2%)	15 (37.5%)	0	0	0	0	n/a	n/a	16 (17.8%)	15 (14.7%)
Partially consistent - WTP values all equal	18 (41.9%)	11 (27.5%)	4 (66.7%)	7 (33.3%)	3 (60.0%)	4 (57.1%)	n/a	n/a	25 (27.8%)	22 (21.6%)
Inconsistent	2 (4.7%)	7 (17.5%)	2 (33.3%)	6 (28.6%)	0	0	7 (19.4%)	8 (23.5%)	11 (12.2%)	21 (20.6%)
Column totals	43 (100%)	40 (100%)	6 (100%)	21 (100%)	5 (100%)	7 (100%)	36 (100%)	34 (100%)	90 (100%)	102 (100%)

Tests of differences in the distribution of consistent responses:

Rank 123 ($\chi^2 = 4.066$, $p=0.254$); Rank 113 ($\chi^2 = 3.211$, $p=0.201$); Rank 122 ($\chi^2 = 0.110$, $p=0.740$); Rank 111 ($\chi^2 = 0.173$, $p=0.677$);

All ranks ($\chi^2 = 3.050$, $p=0.384$)

Table 8 Variable list

Variable name	Definition
AGE	Age of respondent in years
SEX	Sex of respondent: 1=female; 0=male
EDUC	Education level: 1=leaving certificate/third level/diploma; 0=none/primary/junior certificate/intermediate certificate
HEALTH	Respondents' self-assessed health: 1=good/very good; 0=otherwise
INC_A	Gross annual income: 1=£5,000 - £14,999; 0=otherwise
INC_B	Gross annual income: 1=£15,000 - £24,999; 0=otherwise
INC_C	Gross annual income: 1=£25,000 - £34,999; 0=otherwise
INC_D	Gross annual income: 1=£35,000 - £44,999; 0=otherwise
INC_E	Gross annual income: 1=£45,000 and above; 0=otherwise
RISK	Respondents' self-perceived risk of ever needing to use any the programmes: 1=above average; 0=otherwise
EXP	Whether respondents have had first hand experience of cancer, heart disease or community care services: 1=yes; 0=no
DIFFIC	Whether respondents found it difficult to answer the WTP questions: 1=yes; 0=no
INTEREST	Degree of interest shown by respondents as judged by the interviewer: range 1-5 where 1=not at all interested and 5=extremely interested

Table 9 OLS regressions for each of the three programmes

Variable	Ln (WTPC)		Ln (WTPH)		Ln (WTPCC)	
	Std	Incr	Std	Incr	Std	Incr
CONST	-0.948 (-0.835)	2.817* (1.874)	-0.965 (-0.766)	2.501 (1.627)	-1.655 (-1.433)	1.014 (0.694)
AGE	-0.003 (-0.227)	-0.024* (-1.831)	-0.004 (-0.308)	-0.034** (-2.582)	0.011 (0.848)	-0.027** (-2.086)
SEX	0.001 (0.002)	0.411 (1.042)	0.032 (0.085)	0.237 (0.628)	0.308 (0.912)	0.309 (0.863)
EDUC	0.084 (0.219)	0.111 (0.217)	0.410 (0.905)	0.078 (0.151)	0.251 (0.634)	0.711 (1.526)
HEALTH	0.687 (1.353)	0.704 (1.422)	0.395 (0.716)	0.679 (1.284)	-0.020 (-0.035)	0.951* (1.838)
INC_A	0.188 (0.343)	-0.694 (-0.863)	0.139 (0.220)	0.837 (0.958)	0.770 (1.392)	1.214 (1.525)
INC_B	0.382 (0.588)	-0.004 (-0.005)	0.327 (0.427)	1.556* (1.736)	1.122* (1.715)	1.697** (2.018)
INC_C	0.618 (0.885)	-0.525 (-0.431)	0.638 (0.763)	1.062 (0.809)	1.163 (1.652)	1.163 (0.996)
INC_D	0.992 (1.125)	0.830 (0.778)	1.112 (1.171)	2.061* (1.869)	1.491* (1.743)	2.321** (2.329)
INC_E	2.157 (1.391)	0.990 (0.620)	-0.684 (-0.385)	2.775* (1.680)	0.540 (0.364)	3.345** (2.260)
RISK [#]	-0.007 (-0.013)	0.602 (0.451)	1.132** (2.229)	-0.342 (-0.648)	-1.009* (-1.752)	0.646 (1.205)
EXP [#]	0.278 (0.745)	0.393 (0.922)	-0.350 (-0.821)	-0.004 (-0.010)	0.607 (1.332)	0.914** (2.101)
DIFFIC	0.173 (0.345)	0.413 (-0.359)	0.599 (1.010)	0.081 (0.195)	0.829 (1.598)	0.485 (1.274)
INTEREST	0.744*** (3.915)	0.248 (1.158)	0.758*** (3.748)	0.116 (0.550)	0.660*** (3.532)	0.063 (0.302)
n	76	55	74	56	70	55
Adjusted R ²	0.278	0.286	0.259	0.327	0.330	0.389
<i>Misspecification tests (p values):</i>						
Ramsey	0.696	0.809	0.654	0.298	0.902	0.102
Cook-Weisberg	0.419	0.039**	0.987	0.144	0.851	0.483

Notes: Figures in parentheses are t-statistics. *p ≤ 0.10, **p ≤ 0.05, ***p ≤ 0.01

[#] Refers only to the programme which features in the regression equation

References

Arrow K., Solow R., Portney P., Leamer E., Radner R. and Schuman, H. (1993). Report of the NOAA Panel on Contingent Valuation. *Federal Register*, **58**: 4601-4614, Washington DC.

Bishai D.M. and Lang H.C. (2000). The willingness to pay for wait reduction: the disutility of queues for cataract surgery in Canada, Denmark and Spain. *Journal of Health Economics*, 19, 219-230.

Blumenschein K. and Johannesson M. (1998). The relationship between quality of life instruments, health state utilities and willingness to pay in patients with asthma. *Annals of Allergy, Asthma and Immunology*, 80(2), 189-194.

Chiu L., Tang K-Y., Shyu W-C. and Chang T-P. (1999). The willingness of families caring for victims of stroke to pay for in-home respite care: results of a pilot study in Taiwan. *Health Policy*, 46, 239-254.

Davey P., Grainger D., MacMillan J., Rajan N., Aristides M. and Dobson M. (1998). Economic evaluation of insulin lipro versus Neutral (regular) insulin therapy using a willingness to pay approach. *Pharmacoeconomics*, 13(3), 347-358.

Diener A., O'Brien B. and Gafni A. (1998). Health care contingent valuation studies: a review and classification of the literature. *Health Economics*, 7(4), 313-326.

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., Hundley V. and Mapp T. (1998). Willingness to pay: a method for measuring preferences for maternity care? *Birth*, 25, 33-40.

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, 145-159.

Donaldson C., Kidholm K., Olsen J.A. and Shackley P. (2000). Eliciting ex ante willingness to pay for publicly-funded goods: a comparison of methods. Paper presented at the final EuroWill meeting, Dublin, May 2000.

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

Donaldson C., Mapp T., Farrar S., Walker A. and Macphee S. (1997b) Assessing community values in health care: is the "willingness to pay" method feasible? *Health Care Analysis*, 5: 7-29.

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

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: 1063-1070.

Johnson F.R., Desvovages W.H., Ruby M.C., Stieb D. and De Civita P. (1998). Eliciting health state preferences: an application to willingness to pay for longevity. *Medical Decision Making*, 18 (suppl.), S57-S67.

Jones-Lee M.W. (1989). *The Economics of Safety and Physical Risk*. Basil Blackwell, Oxford.

Kahneman D. and Knetsch J. (1992). Valuing public goods: the purchase of moral satisfaction. *Journal of Environmental Economics and Management*, 22, 57-70.

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

Lee S.J., Lilies B., Gnomon P.J., Winston M.C. and Johannesson M. (1998). The impact of risk information on patients' willingness to pay for autologous blood donation. *Medical Care*, 36(8), 1162-1173.

O'Connor R.M., Johannesson M, Haas S.L. and Kobelt-Nguyen G. (1998). Urge incontinence: quality of life and patients' valuation of symptom reduction. *Pharmacoeconomics*, 14(5), 531-539.

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

Olsen J.A., Pereira J., Donaldson C. and Shackley P. (2000). *The insensitivity of 'willingness to pay' to the size of the good: new evidence for health care*. Paper presented at the final EuroWill meeting, Dublin, May 2000.

Ryan M. (1998). Valuing psychological factors in the provision of assisted reproductive techniques using the economic instrument of willingness to pay. *Journal of Economic Psychology*, 19, 179-204.

Ryan M. and San Miguel F. (2000). Testing for consistency in willingness to pay experiments. *Journal of Economic Psychology*, 21, 305-317.

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

Schkade D. and Payne J. (1994). How people respond to contingent valuation questions: a verbal protocol analysis of willingness to pay for an environmental regulation. *Journal of Environmental Economics and Management*, 26, 88-109.

Schoemaker P.J.H. (1982) The expected utility model: its variants, purposes, evidence and limitations. *Journal of Economic Literature*, 20: 529-563.

Shackley P. and Dixon S. (2000). Using contingent valuation 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* (forthcoming).

Zethraeus N. (1998). Willingness to pay for hormone replacement therapy. *Health Economics*, 7, 31-38.

Appendix Programme descriptions

Pain treatment for cancer patients

200 more patients with advanced cancer could have relief from pain by radiotherapy in addition to the 1,600 who are currently getting this treatment.

Without this treatment they would get pain-reducing medicine. Many patients will still not have satisfactory pain relief, while others will get significant side effects in the form of tiredness and poor quality of life.

Radiotherapy for these patient groups may have good pain relieving effects among 75% and lead to improved functioning among most patients. The treatment will have few side effects. On average patients will benefit from this treatment in their last year of life. the treatment will not prolong patients' lives.

There are patients in every age group and the average is 60 years old. Men and women are affected in equal numbers.

Heart operations

100 more heart operations can be provided each year in addition to the 600 which are currently done in the country.

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

The operation will make 75% of patients completely free from pain, with less pain for the rest. Without the operation the patients will expect to live 8-10 years. With the operation they will on average live for an extra year on top of this.

The operation mortality risk is 1% (so 1 in 100 people will die whilst being operated on).

Community care

200 more physically and mentally dependent elderly people would be able to remain in their own homes as a result of an expansion in community care services, thereby reducing the current admissions to long-stay care from the present level of 6,000 per year,

The additional community care services would be in the areas of home nursing, home help and day care facilities. The additional services would be targeted at highly dependent elderly people living at home.

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

The majority of people benefiting from this programme will be women aged 75 years and over.