



ELSEVIER

Journal of Health Economics 21 (2002) 585–599

JOURNAL OF
**HEALTH
ECONOMICS**

www.elsevier.com/locate/econbase

Do ordering effects matter in willingness-to-pay studies of health care?

Jennifer M. Stewart^{a,*}, Eamon O'Shea^b, Cam Donaldson^c,
Phil Shackley^d

^a *Centre for Studies of Children at Risk, Department of Psychiatry and Behavioural Sciences,
McMaster University, Hamilton, Ont., Canada*

^b *Department of Economics, National University of Ireland, Galway, Ireland*

^c *Departments of Economics and Community Health Sciences, and Centre for Health and Policy Studies,
University of Calgary, Calgary, Alta., Canada*

^d *Sheffield Health Economics Group, University of Sheffield, Sheffield, UK*

Received 11 January 2001; received in revised form 25 July 2001; accepted 11 December 2001

Abstract

Willingness-to-pay studies are increasingly being used in the evaluation of health care programmes. However, there are methodological issues that need to be resolved before the potential of willingness-to-pay can be fully exploited as a tool for the economic evaluation of health care programmes. Of particular methodological interest are the consequences of varying the order in which willingness-to-pay questions are presented to respondents in contingent valuation studies. This paper examines the possibility of ordering effects in willingness-to-pay studies in health care. That is, when asking willingness-to-pay questions about three health care programmes within a single survey, does the order the programmes are presented have an impact on the reported willingness-to-pay? Ordering effects are observed in the ranking of the programmes, in the proportion of zero values reported and in the WTP for one of the programmes. The results suggest that the best explanation for the ordering effects is one of fading glow, whereby the first programme in any sequence captures much of the utility associated with giving. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: D61; I11; I18

Keywords: Willingness-to-pay; Ordering effects; Survey design

* Corresponding author. Present address: Department of Economics, McMaster University, Hamilton, Ont., Canada L8S 4M4. Tel.: +1-905-525-9140x23672; fax: +1-905-521-8232.

E-mail address: jstewar@mcmail.cis.mcmaster.ca (J.M. Stewart).

1. Introduction

Research on eliciting willingness-to-pay (WTP) values in order to evaluate projects has been extensive in environmental economics and is growing in health economics (Diener et al., 1998; Klose, 1999; Smith, 2000). In 1993, the National Oceanic and Atmospheric Administration (NOAA) published guidelines for the use of contingent valuation (CV) studies of environmental resource damage. However, the publication of these guidelines has not ended the debate about the 'best' survey design, particularly in the health care field. For that reason, research is ongoing into the reliability of CV studies and most of the methodological issues raised in the environmental field remain when CV studies are used in health care.

This paper explores the possibility of ordering effects in CV studies of health care programmes as part of a larger research project (EuroWill)¹ to examine CV studies in the health care field. This paper asks, "does the WTP for a programme change when the programme is in a different position in a sequence of programmes?" This question is answered by surveying two samples and reversing the order of the programmes between the two samples. This paper builds on previous, methodologically diverse, research on ordering effects in the environmental field (Gorden, 1969; Brookshire et al., 1981; Boyle et al., 1990, 1993; Bateman and Langford, 1997) and in the health care field (Kartmann et al., 1996; Halvorsen, 1996). The results from previous literature have been mixed, with support both for and against ordering effects. However, this literature is sparse with only a small number of published studies having examined whether an ordering effect arises when CV questions are asked of multiple programmes in a single survey instrument.

2. Conceptual issues and hypotheses

Why elicit WTP for several programmes within one survey? There are three reasons. First, the NOAA guidelines recommend that respondents be reminded of substitutes when reporting WTP values because, "if individuals fail to consider seriously the public or private goods that might be substitutes for the resources in question, their responses to questions in a CV survey may be unrealistically large" (Arrow et al., 1993). Second, if WTP values are to be used to aid resource allocation then values are needed for several programmes. Third, respondents may view a single programme as representing the wider health sector to which they have a positive attitude and overestimate their valuation of a single programme. By asking about several programmes in one survey, the potential for overestimation is reduced. But including several programmes may introduce other biases, especially ordering effects, which we examine in this paper.

We conducted a study that elicited WTP for three programmes. These three programmes were presented in reverse order to two samples. Appendix A contains the complete description of the programmes which were: an increase in pain-relieving treatments for cancer patients (C), an increase in heart operations (H), and an increase in community care services

¹ EuroWill, a project in which various methodological issues, arising from the use of WTP to help set priorities across competing uses of health care resources, were addressed through surveys in six European countries.

(CC). The order in the first sample was C_1^1 , H_1^2 , and CC_1^3 and in the second sample, CC_2^1 , H_2^2 , and C_2^3 , where the superscripts refer to the order of the programme within the survey and the subscripts refer to the sample. The WTP questions were put in the form of an either/or sequence rather than an additional good sequence. Respondents were informed that the programmes were in competition with each other for funding, but they were explicitly asked to ignore the other two programmes when providing a WTP value for any one programme.

For either/or sequences, traditional economic theory does not provide an explanation for ordering effects. Under rational choice theory, preferences, expressed in terms of WTP, should not depend upon the order in which options are presented to respondents. An ordering effect violates the principle of procedural invariance of rational choice theory. Therefore, either framing effects or other elements in the respondent's utility function² are at work or respondents perceive the sequence as an additional good sequence. For additional good sequences, economic theory does provide an explanation for ordering effects in the form of income and substitution effects that may occur when a list of possibilities is expanded (Carson and Mitchell, 1995; Carson et al., 1998). Thus, the first hypothesis we tested was whether respondents perceived the sequence as an either/or or additional good sequence.

Hypothesis 1. Either/or sequence versus additional goods: respondents view the sequence of goods as an either/or sequence rather than as an additional good sequence.

Respondents were asked to treat the three programmes as substitutes for each other. Our expectation was that respondents would not add up their WTP values. However, there was no guarantee that respondents would act in this way, particularly because they were also reminded to consider that their contribution would reduce what they had left to spend on other things. We predict that if respondents did add up their WTP values then, as the survey progressed, respondents may have become aware that they had less to spend after they had already reported contributions for other programmes. Therefore, the reported WTP for a programme may be lower if it was last in the sequence than if it was first in the sequence. Specifically: $C_1^1 > C_2^3$, $CC_2^1 > CC_1^3$, and $H_1^2 = H_2^2$.

We next consider two hypotheses as to why the order of the programmes may have an impact on the reported WTP in an either/or sequence.

Hypothesis 2. Starting point bias: respondents require a framework in order to answer WTP questions and latch on to recently heard or said values.

A starting point bias is a framing effect where respondents are influenced by the first numbers presented. We extend this idea to include any series of numerical responses and numbers supplied by the respondents themselves. In our survey, the most preferred programme would elicit a high WTP value when placed at the beginning of the sequence of questions. As respondents answer the subsequent questions, they latch on to this first value and their WTP values for the other programmes will be biased upwards. If the ranking of programmes is such that the cancer programme is preferred to the community care programme,

² See Kahneman and Tversky (1984), Tversky et al. (1988) for the seminal work in the area of individual preferences and decision making from a psychological viewpoint applied to economics. Dolan (1997) provides a framework for discussing recent economic work on individual preferences.

regardless of the order of the programmes in the sequence, then a starting point bias occurs if the WTP valuation yields the following result: $C_1^1 > C_2^3$, $CC_1^3 > CC_2^1$, and $H_1^2 > H_2^2$.

We can also examine the number of zero WTP values for each programme to test our hypothesis. If a starting point bias occurs then, when the least preferred programme is presented first, we expect that the downward bias will lead to more reported zero values.

Hypothesis 3. Fading glow: respondents receive moral satisfaction or warm glow from simply contributing to a publicly financed good and the first programme in the sequence is likely to capture the bulk of this moral satisfaction.

Respondents receive moral satisfaction or a warm glow from contributing to public goods.³ It is likely, therefore, that they will receive a warm glow from saying they would be willing to contribute to a health care programme. The health care programme is valued partly for the moral satisfaction associated with the contribution in addition to being an end in itself. Assuming that each of the three programmes carries the same potential for moral satisfaction, the first programme in the sequence is the crucial programme because it provides the initial opportunity for respondents to demonstrate their willingness to help, thereby exhausting the bulk of the warm glow associated with the decision to support a good cause. The bulk of the warm glow is captured by the WTP for the first programme and the warm glow element of the WTP for subsequent programmes is less, hence, fading glow. A fading glow bias would yield the result: $C_1^1 > C_2^3$, $CC_2^1 > CC_1^3$, and $H_1^2 = H_2^2$. We also predict that when a programme is last in a sequence it will have more zero WTP values than when it is presented first.

Unfortunately, the predicted pattern of the WTP values is the same for hypotheses 1 and 3. If we assume that the rate at which warm glow fades is not related to income and that the probability of hitting a budget constraint is inversely related to income, then we may be able to distinguish between these two hypotheses. We expect WTP to increase with income, but if the income elasticity is less than one, on the assumption that health care is a normal good, then respondents with higher incomes will be less likely to hit their budget constraint. The vast majority of within-country studies on income elasticities for health care programmes report values of less than one, suggesting that people commonly perceive health care to be a necessity (Folland et al., 2001). If respondents do perceive the questions as an additional goods sequence then we predict that respondents with higher incomes are not influenced by the order of the questions to the same degree as respondents with lower incomes. We test this hypothesis by running regressions that include an interaction term between sample and income.

We also considered one final hypothesis about ordering effects and the respondent's previous experience with the programme.

Hypothesis 4. Previous experience: respondents with previous experience with the programme have better knowledge of its value to them and will not be influenced by the order of the WTP questions.

³ See Kahneman and Knetsch (1992), Andreoni (1990) for the theoretical exposition of moral satisfaction and warm glow and Chilton and Hutchinson (2000) for empirical evidence in CV studies.

Previous research indicates that respondents with experience of the good are not influenced by the order of the WTP questions (Boyle et al., 1993; Kartmann et al., 1996). Respondents with experience may have better knowledge of the value of the programme to them and are not influenced by the order of the questions. For those with previous experience with the programme, we predict that there will be no difference in the WTP values between the samples, i.e. $C_1^1 = C_2^3$, $CC_1^3 = CC_2^1$, and $H_1^2 = H_2^2$. For those without previous experience, we predict that there may be a difference in the WTP values between the samples, i.e. $C_1^1 \neq C_2^3$, $CC_1^3 \neq CC_2^1$, and $H_1^2 \neq H_2^2$. We divided our sample into groups by whether or not they had previous experience with the health condition. These groups are not mutually exclusive in that a respondent may have had previous experience with more than one health care programme or condition.

3. Methods

3.1. Sampling

The survey was carried out in the largely rural Western Health Board region of Ireland, which contains a population of approximately 350,000 people. The sample design was based on a two-stage clustered sample using the Electoral Register as a population frame. The data generated from the electoral register was re-weighted on the basis of the principal economic status of head of household, household composition and sub-regional classification to make it representative of the overall population in the region. A total of 473 people were selected for interview. The overall response rate was 45%. This rate is less than that achieved in some surveys in which WTP has been used in interviews of the general public, but more than in other such surveys (Donaldson et al., 1997a; Olsen and Donaldson, 1998). The main difficulty for the interviewers, even with return visits, was meeting the people selected for interview face-to-face. Once contact was made the response rate was high with only 8% of people refusing to be interviewed when met face-to-face by the interviewers. One of the potential problems with using the electoral register as the population frame is that some younger people in Ireland return home to vote but live elsewhere in the country. This trend is particularly true of the region used for the survey, which has seen very high levels of outward migration in recent decades.

3.2. Questionnaire

Respondents were given introductory information outlining the objective of the survey. The three programmes were named and the respondents were asked to think about them “as if they are in competition with each other for funding”. Respondents were then asked about their perception of ever needing the programmes and their past experience with the health states. Next, respondents were presented with the first two programme descriptions and asked to rank these two programmes. Then respondents were presented with the third programme and asked to rank all three programmes. The order of these questions, in terms of the health care programmes to which they referred, was reversed between the two samples to correspond with the order of the WTP questions.

The next section of the survey asked the respondents about their WTP for the programmes. For each programme, this section began by asking respondents if they would be willing to contribute anything in extra taxation for the given expansion. All respondents, regardless of their answer to the taxation question, were then asked if they would be willing to pay if the payment was in terms of a voluntary contribution. The inclusion of the voluntary option is important because it provides a payment option for those people who, for whatever reason, distrust public mechanisms of health care resource allocation and prefer more direct voluntary contributions. If the respondent answered no to both of these questions, they were asked to explain the reasons why they were unwilling to pay. Otherwise, respondents were asked the following question: “How much is the MAXIMUM your household would be willing to contribute each year for this expansion in the (relevant programme)? Please bear in mind that your contribution would reduce what you have left to spend on other things”.

To help them answer the question, respondents were shown the payment card⁴ included in Appendix A. Respondents were then asked to indicate the reason they were willing to contribute to this programme.

The issue of how to interpret a zero WTP is also important, particularly if a significant proportion of the respondents indicate that they are unwilling to contribute to the programme. Previous research (Olsen and Donaldson, 1998) has indicated that a reported WTP of zero does not necessarily indicate a true WTP of zero, but rather some respondents protest against paying for the programmes. Respondents unwilling to contribute in this study were asked to give a reason for that decision. Two of the possible answers were “this programme is of no value to my household” and “I cannot afford it”. If one of these was the reason why the respondent reported a zero WTP then we concluded that zero is the true WTP and we refer to these responses as true zeros. We present all of our regression results for a sample that includes only true zeros, but we did conduct all empirical work on the full sample and the results did not change.

The last section of the questionnaire asked questions on socio-economic, health, and demographic details. Table 1 provides a description of each variable used in our analysis. We ran interval regressions⁵ to control for observed differences between the two samples and to test for the internal consistency of the reported WTP values. We used this approach because the reported WTP values were grouped. All respondents reported a WTP value that was on the payment card and so we interpreted the respondent’s choice as indicating they were at least willing-to-pay the stated amount, but not willing-to-pay the next highest amount. We assumed that the true WTP value was in this range. The interval regression also controlled for any censoring to the WTP values that may have occurred, because the highest WTP value on the payment card was £200. Furthermore, we used the log of the WTP value because its distribution was positively skewed.

⁴ The payment card is a commonly used method in studies of health care. Its validity has been demonstrated against open-ended questions (Donaldson et al., 1997b). Although dichotomous choice questions are also popular, their validity has been questioned recently, due to problems of ‘yea-saying’ (Holmes and Kramer, 1995; Ready et al., 1996; Kramer and Mercer, 1997). In the EuroWill project, another study was set-up to compare payment cards with dichotomous choice questions (Ryan et al., 2000).

⁵ See Donaldson et al. (1998) about the use of discrete variable analysis in WTP studies. We also ran linear OLS, ordered probit, and ordered logit models. The results were similar for all models. We present the interval regression because, it was the most appropriate of the models we estimated.

Table 1
Variable specification

Female	1 for female, 0 for male
Age	Age in years
Age-squared	Age-squared in years
Single	1 for never married/single, 0 for other status
Primary education	1 for highest level of education of a primary certificate, 0 for higher levels
Own health < good	1 for self reported health status of 'neither good nor bad' or 'poor', 0 for 'very good' or 'good'
Smoker	1 for smoke daily, 0 for smoke occasionally or never
Income	Logarithm of the midpoint of the income interval in Irish pounds adjusted for number of persons in the household (OECD weights: 1 for first adult, 0.7 for additional adults, 0.5 for each child)
Sample 2	for sample 2 where the order of the programmes was community care, hearts, and cancer, 0 for sample 1 where the order of the programmes was reversed
Experience	1 if answered yes to "have you or anyone in your close family ever had personal, first hand experience of (the relevant condition)?"

4. Results

Table 2 describes how respondents ranked the programmes in the two samples. The cancer programme was ranked most important by a larger percentage of respondents than the other two programmes in both samples. It was more difficult to decide the ranking between the other two programmes. We concluded that hearts was ranked higher than community care because a larger proportion ranked it second most important.

As Table 2 indicates, there was a difference in the proportions in the rankings between the samples. Using Pearson's χ^2 statistic, we found that there was a statistically significant difference in the ranking of the cancer programme. A higher proportion ranked the cancer programme most important in the first sample, the sample in which respondents were first given the description of the cancer programme. At this early point in the survey we observed ordering effects and in a relatively simple exercise compared to answering WTP questions. The difference in the ranking of the other two programmes between the samples was not statistically significant.

Table 3 presents the unconditional means of the WTP for the three programmes for each sample. The pattern between the two samples is $C_1^1 > C_2^3$, $H_1^2 > H_2^2$, and $CC_2^1 > CC_1^3$ which is consistent with either an additional good sequence or a fading glow hypothesis. However,

Table 2
Ranking of programmes by sample^a in percentages (number of respondents in parentheses)

Sample	Cancer		Heart		Community care	
	1	2	1	2	1	2
Most important	79.65 (90)	63.67 (64)	47.79 (54)	50.50 (51)	52.21 (59)	52.48 (53)
Second most important	17.79 (20)	26.73 (27)	35.40 (40)	31.68 (32)	11.50 (13)	14.85 (15)
Least important	2.65 (3)	9.90 (10)	16.81 (19)	17.82 (18)	36.28 (41)	32.67 (33)
Pearson's χ^2	8.55		0.33		0.66	

^aOrder in sample 1 cancer, hearts, community care. Order in sample 2 community care, hearts, cancer.

Table 3
Descriptive statistics of WTP by sample^a

	Cancer	Heart	Community care
Sample 1			
Mean	49.22	42.37	37.14
S.D.	61.56	56.36	54.54
Median	20.00	20.00	10.00
BS ^b 95% CI	38.08–62.65	32.34–54.64	27.34–48.88
Zeros	11	15	13
<i>n</i>	102	99	96
Sample 2			
Mean WTP	41.40	37.13	42.12
S.D.	51.81	48.04	51.41
Median	20.00	20.00	20.00
BS ^b 95% CI	31.26–52.22	28.15–49.29	32.05–53.95
Zeros	5	4	2
<i>n</i>	86	80	91

^aOrder in sample 1 cancer, hearts, community care. Order in sample 2 community care, hearts, cancer.

^bBootstrap.

using the bootstrap confidence interval,⁶ we could not reject the null hypothesis that the mean WTP for a programme in the first sample was equal to its mean WTP in the second sample for all three programmes, i.e. $\widehat{H}_0 : C_1^1 = C_2^3, H_1^2 = H_2^2$, and $CC_1^3 = CC_2^1$. Similarly, we could not reject the null hypothesis that the mean WTPs of the three programmes were equal, i.e. $\widehat{H}_0 : C_1^1 = H_1^2 = CC_1^3$ and $C_2^3 = H_2^2 = CC_2^1$.

Table 4 contains the results from the interval regressions and the observed relationship was $C_1^1 > C_2^3$ and $CC_2^1 > CC_1^3$. The higher WTP for community care in sample two was statistically significant. For the other two programmes, the difference between the samples was not statistically significant. The fading glow hypothesis and the additional good sequence both predict this observed difference for community care. To distinguish between these two hypotheses, we predict that if respondents perceived the sequence as an additional good sequence then we would observe that question order was correlated with income. However, this prediction is only plausible if the income elasticity was less than one. The coefficients in Table 4 show that the income elasticity was between 0.61 and 0.69. When we ran regressions that included an interaction term between sample and income, we found that the interaction term was not statistically significant and the coefficient for the sample variable did not change. We concluded that question order does not have a differential impact by income. This result gives some support for assuming that respondents were viewing the WTP questions as either/or questions as intended, leaving fading glow as the most plausible explanation for the observed difference for community care. We do not present the results here from these regressions, because the coefficients on the other independent variables were similar to those in Table 4.

Accepting both the fading glow hypothesis and that respondents viewed the questions as an either/or sequence is consistent. We assume that warm glow is independent of the

⁶ See Efron and Tibshirani (1993) for a detailed discussion of the use of bootstrapping.

Table 4
Interval regression of log of WTP (standard errors in parentheses)

	Cancer	Heart	Community care
Female	-0.3024 (0.2304)	-0.3441 (0.2460)	0.1199 (0.2244)
Age	0.0425 (0.0435)	0.03436 (0.0472)	0.0259 (0.0419)
Age-squared	-0.0005 (0.0003)	-0.0004 (0.00004)	-0.0004 (0.0004)
Single	-0.6493** (0.3169)	-0.5480* (0.3218)	-0.3261 (0.3012)
Primary education	-0.7752** (0.3188)	-0.9466** (0.3336)	-0.4695 (0.3121)
Own health < good	0.1635 (0.3034)	0.2617 (0.3159)	0.1224 (0.2951)
Smoker	0.2145 (0.2908)	0.2104 (0.3039)	0.0548 (0.2871)
Income	0.6103** (0.2343)	0.6814** (0.2612)	0.6615** (0.2282)
Sample 2	0.0285 (0.2262)	0.0484 (0.2376)	0.4078* (0.2185)
Constant	3.6298** (0.2549)	3.5353** (0.2755)	2.9597** (0.2500)
Sigma	1.3505** (0.0882)	1.3891** (0.0928)	1.2938** (0.0832)
log-likelihood	-417.3773	-397.8845	-409.4389
Null log-likelihood	-443.8032	-421.2037	-428.8905
Likelihood ratio index	0.0595	0.0554	0.0454
Sample size	158	151	154
Number of zeros	16	19	15

Significant at (*) 90% CL; (**) 95% CL. The baseline characteristics are male with the mean age, previously married, more than a primary education, very good/good health, non-smoker, mean income, and in sample 1.

programme being evaluated and is captured by the first programme. Once respondents indicate a willingness to contribute, they do not find it necessary to include that value in their subsequent WTP values. Respondents are evaluating the programmes in isolation and not adding up the values, but do not include warm glow after it has been expressed once. However, it would require a high degree of cognitive compartmentalisation among respondents to be able to rule out fully any relationship among the three programmes.

Table 5 shows the proportion of the respondents that reported a zero WTP for each programme by sample. It is apparent that there were fewer true zeros in the second sample. Using Pearson's χ^2 statistic, we found that the difference in the proportion of zeros reported for the heart and community care programmes was statistically significant, but for the cancer programme the difference was not statistically significant. This result is also consistent with a fading glow hypothesis.

Our fourth hypothesis is that the ordering effects would not occur for respondents with previous experience. We divided our sample into two sub-samples; those with previous experience with cancer and those without. We then regressed the log of WTP on the independent variables for each of these sub-samples separately. We repeated this procedure for

Table 5
Proportion of sample reporting a zero WTP

Sample	Cancer		Heart		Community care	
	1	2	1	2	1	2
Proportion of sample	9.73	4.95	13.27	3.96	11.50	1.98
Number of zeros	11	5	15	4	13	2
Pearson's χ^2	1.76		5.72		7.42	

Table 6

Impact of question order on log of WTP for each programme by experience with the condition (standard errors in parentheses)

Experience with condition	WTP for programme		
	Cancer	Heart	Community care
Experience with cancer	-0.6234 (0.4143)	-0.4430 (0.4244)	0.0570 (0.3762)
No experience with cancer	0.4004* (0.2250)	0.1897 (0.2367)	0.6069** (0.2343)
Experience with hearts	-0.2735 (0.3815)	-0.1050 (0.3820)	0.3654 (0.3470)
No experience with hearts	0.2234 (0.2676)	0.2459 (0.2889)	0.5080* (0.2674)
Experience with community care	-1.2884** (0.6149)	-1.0493** (0.4426)	-0.4208 (0.4938)
No experience with community care	0.1912 (0.2495)	0.2462 (0.2782)	0.5942** (0.2515)

Significant at (*) 90% CL; (**) 95% CL. The baseline characteristics are male with the mean age, previously married, more than a primary education, very good/good health, non-smoker, mean income, and in sample 1. Reported coefficient is for sample 2.

the heart and the community care programmes. Table 6 reports the coefficients from the sample variable, which indicates the impact of the question order on the WTP values.

We found, as expected, that significant programme specific ordering effects did not occur for respondents with previous experience with cancer, hearts and community care. Conversely, as expected, programme specific ordering effects occurred for respondents without experience with cancer and community care. There was evidence, however, of some interaction between the three programmes with respect to previous experience with community care. Respondents with experience with community care had a significantly lower WTP for the cancer programme and the heart programme when community care was presented first in the sequence.

5. Discussion

This paper confirms the existence of ordering effects in CV studies in health care. Asking people to put a monetary WTP value on a sequence of health care programmes will likely yield different results depending on the order of the programmes because answers to earlier questions will affect responses to later questions. We found that the fading glow hypothesis best explains the observed ordering effects. The fading glow refers to the utility respondents gain from saying that they are willing to contribute to the health care sector rather than the impact of adding additional goods to a sequence. The ordering effect arises from the tendency of respondents, regardless of income, to overstate their WTP for the first programme in a sequence. Subsequent valuations will be reduced because the first programme picks up the major benefits to the respondent from making a contribution.

For further evidence of this effect, we examined the reasons why respondents indicated that they were willing-to-pay for the various programmes. The responses can be divided into four categories.

1. Selfish: "I, or a member of my household, might benefit", and "a member of my family/friend has used the service".

2. Altruistic: “other people will benefit”, “more people will be able to return to their normal activities”, “the programme will improve health”, and “more equal access to health care”.
3. Existence value: “reassuring to know care is available”.
4. Technical: “I would support technical progress in medicine”.

For community care, we found that the difference in the distribution of the responses across these categories between the two samples was statistically significant. In sample 2, there were more altruistic responses and fewer selfish responses, which is consistent with the fading glow hypothesis. The pattern was the opposite for cancer, which again is consistent with the fading glow hypothesis, but the difference here was not statistically significant.

The presence of fading glow is not sufficient to generate a statistical difference in WTP between the two samples for the cancer programme. A probable explanation is that the strong preference for the cancer programme among respondents is dominating the fading glow effect. Further support for this explanation is the lack of difference in the number of zeros between the two samples for the cancer programme and the lack of difference in the reasons given for WTP between the two groups for the cancer programme. Whether higher ranked programmes will always dominate fading glow effects is an empirical question which can only be answered through careful analysis of both initial rankings and subsequent WTP estimates. Ordering effects will affect WTP estimates but the degree of influence will be determined by the strength of consumer preference for the programme ranked highest in initial rankings and the extent of fading glow in the decision-making calculus of respondents.

There may also be other type of interaction among the programmes. When we examined the WTP by experience with the health condition, the results suggest that there are complex and not always easily understood interactions among health care programmes. When CC was presented first, respondents who have previous experience with CC reported lower WTP for the other two programmes. A possible explanation is that asking people to first value CC reminds them of the high value attached to remaining in your own home relative to any type of in-patient care. Consequently, they will ascribe a lower WTP to the remaining two programmes, neither of which involve long-stay care, but both of which require hospitalisation and potentially invasive health care intervention. It is impossible to be precise here, but some form of trade-off is occurring with people making valuation judgements on the basis of their relative experience with different types of care.

These results raise some concerns over the use of CV in health care. If the rankings of the programmes are affected by the order in which they are presented to respondents then it is difficult to be sanguine about the validity of the various WTP values elicited by the process. Policy makers should be aware that when respondents are asked to evaluate multiple programmes even the rankings are likely to be influenced by the question order. Failure to recognize this issue will lead to spurious decision making in the allocation of scarce health care resources. The fading glow hypothesis states that people use up much of their goodwill for health care in providing a WTP for the first programme they are asked to value, therefore, order matters, particularly if there are not strong rankings among the programmes. Evaluating a programme in isolation will produce an overestimate of the WTP value because of the existence of warm glow. However, simply adding more programmes

and measuring the difference in the WTP among programmes will not necessarily produce an estimate of warm glow because one of the programmes may have a higher potential to satisfy the moral satisfaction element of the valuation process for public goods. If programmes differ in terms of their capacity to deliver moral satisfaction for respondents then the positioning of the programme in the valuation sequence will affect WTP estimates for all programmes. This possibility points to the need to be careful both in the selection of programmes for valuation and in providing descriptions of the programmes to respondents. It may be that as well as asking respondents prior questions regarding their experience of various programmes under consideration that questions also need to be asked about the moral worth of the programmes to respondents.

Previous experience with the programmes further complicates the estimation process. Previous experience with a health care programme impacts on the valuation of that programme and, crucially, on other programmes. If these problems are to be overcome in CV studies in health care, careful thought needs to be given to what programmes to include in any one survey and whom to survey. Although, restricting and manipulating both the programmes included in the survey and the respondents surveyed may weaken the validity of the results for decision making.

The survey limitations of the present work should be explicitly recognised. In order to untangle further the impact of order and interaction with other programmes a more complicated experimental design would be preferred. This paper considered only two of the six possible orderings of the programmes. Ideally, six versions of survey should be created, one for each possible combination of the three programmes. Similar disaggregation would be required to test for the impact of previous experience on the WTP estimates. A more expansive survey design would, of course, have resource implications, given the number of respondents in each group necessary to allow meaningful comparison and the time-intensive nature of the interviewing.

6. Conclusion

This paper has examined how people's WTP for three health care programmes is related to the order in which the programmes are presented. If WTP estimates are to be used to inform policy decisions, then the degree of bias associated with the estimation procedure needs to be made explicit. If the value of a subsequent programme is influenced by the value given to a previous one then an important axiom, the independence of irrelevant alternatives, is being violated.

Ordering effects are observed in the ranking of the programmes, in the proportion of zero values reported and in the WTP for one of the programmes (CC). The observed pattern of these results suggests that the best explanation for the ordering effects is one of fading glow, whereby the first programme in any sequence captures much of the utility associated with giving. Respondents may feel that they have met their social obligations once they have contributed to the first programme on the list. Previous experience may also matter, both for the WTP valuation of the programme for which the respondent has experience and for the valuation of other programmes. These results raise questions about the use of CV as an aid to resource allocation in health care.

Acknowledgements

The authors would like to thank Ian Bateman, Graham Loomes, Martin Dooley, David Feeny, participants at the EuroWill Dublin Conference, and participants at the Irish Economic Association Meetings, May 2000. The authors would like to thank James Williams and the Economic and Social Research Institute for assistance with the survey. Cam Donaldson holds the Svare Chair in Health Economics and is an Alberta Heritage Senior Scholar, Canadian Institutes of Health Research Senior Investigator and Fellow of the Institute of Health Economics. The EuroWill project was supported by a grant from the European Commission BIOMED 2 Research Programme (PL 950832). The views expressed in this paper are those of the authors and not the funding agency.

Appendix A. Description of programmes

A.1. Pain treatment for cancer patients

Two hundred more patients with advanced cancer could have pain relief from pain by radiotherapy in addition to the 1600 who are currently getting this treatment. Without this treatment they would get pain-reducing medicine. Many patients will 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 the patients' lives. There are patients in every age group and the average age is 60-year-old. Men and women are affected in equal numbers.

A.2. Heart operations

One hundred 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 one in 100 people will die whilst being operated on).

A.3. Community care

Two hundred more physically and mentally dependent elderly people would be able to remain in their own homes as a result of an expansion of community care services, thereby reducing the current admissions to long-stay care from the present level of 6000 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 the people benefiting from this programme will be women aged 75 years and over.

A.4. Payment card

Amount (£)		
0.00	25.00	120.00
2.50	30.00	140.00
5.00	40.00	160.00
7.50	50.00	180.00
10.00	60.00	200.00
15.00	80.00	More than 200.00 (please specify)
20.00	100.00	

In the interview please tick the amounts you are sure you *would pay*; cross beside the amounts you are sure you *would not pay*; circle around the amount which is the *maximum* you would be willing to pay.

References

- Andreoni, J., 1990. Impure altruism and donations to public goods: a theory of warm-glow giving. *Economic Journal* 100, 464–477.
- Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R., Schuman, H., 1993. Report of the National Oceanic and Atmospheric Administration Panel of Contingent Valuation, Washington DC. *Federal Register* 58 (10), 4601–4614.
- Bateman, I.J., Langford, I.H., 1997. Budget constraint, temporal and ordering effects in contingent valuation studies. *Environment and Planning* 29 (7), 1215–1228.
- Boyle, K.J., Reiling, S.D., Phillips, M.L., 1990. Species substitution and question sequencing in contingent valuation surveys evaluating the hunting of several species of wildlife. *Leisure Science* 12, 103–108.
- Boyle, K.J., Welsh, M.P., Bishop, R.C., 1993. The role of question order and respondent experience in contingent-valuation studies. *Journal of Environmental Economics and Management* 25, S80–S90.
- Brookshire, D.S., d'Arge, R.C., Schulze, W.D., Thayer, M.A., 1981. Experiments in valuing public goods. In: Smith, V.K. (Ed.), *Advances in Applied Microeconomics*, Vol. 1. JAI Press, Greenwich, CT.
- Carson, R.T., Mitchell, R.C., 1995. Sequencing and nesting in contingent valuation studies. *Journal of Environmental Economics and Management* 28, 155–173.
- Carson, R.T., Flores, N.E., Hanemann, W.M., 1998. Sequencing and valuing public goods. *Journal of Environmental Economics and Management* 36, 314–323.
- Chilton, S.M., Hutchinson, W.G., 2000. A note on the warm glow of giving and scope sensitivity in contingent valuation studies. *Journal of Economic Psychology* 21, 343–349.
- 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.
- Dolan, P., 1997. The nature of individual preferences: a prologue to Johannesson, Jonsson and Karlsson. *Health Economics* 6, 91–93.
- Donaldson, C., Mapp, T., Farrar, S., Walker, A., Macphee, S., 1997a. Assessing community values in health care: is the 'willingness to pay' method feasible? *Health Care Analysis* 5, 7–29.

- Donaldson, C., Thomas, R., Torgerson, D.J., 1997b. Validity of open-ended and payment scale approaches to eliciting willingness-to-pay. *Applied Economics* 29, 79–84.
- Donaldson, C., Jones, A.M., Mapp, T.J., Olsen, J.A., 1998. Limited dependent variables in willingness-to-pay studies: applications in health care. *Applied Economics* 30, 667–677.
- Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman & Hall, London.
- Folland, S., Goodman, A.C., Stano, M., 2001. *The Economics of Health and Health Care*, 3rd Edition. Prentice-Hall, Upper Saddle River, NJ.
- Gorden, R., 1969. *Interviewing: Strategy, Technique and Choices*. Dorsey, Homewood, IL.
- Halvorsen, B., 1996. Ordering effects in contingent valuation surveys: willingness-to-pay for reduced health damage from air pollution. *Environmental and Resource Economics* 8, 85–499.
- Holmes, T.P., Kramer, R.A., 1995. An independent sample of yea-saying and starting point bias in dichotomous-choice contingent valuation. *Journal of Environmental Economics and Management* 29, 121–132.
- Kahneman, D., Tversky, A., 1984. Choices, values and frames. *American Psychology* 39, 341–350.
- Kahneman, D., Knetsch, J.L., 1992. Valuing public goods: the purchase of moral satisfaction. *Journal of Environmental Economics and Management* 22, 57–70.
- Kartmann, B., Nils-Olov, S., Magnus, J., 1996. Valuation of health changes with the contingent valuation method: a test of scope and question order effects. *Health Economics* 5, 531–541.
- Klose, T., 1999. The contingent valuation method in health care. *Health Economics* 47, 97–123.
- Kramer, R.A., Mercer, D.E., 1997. Valuing a global environmental good: US residents' willingness to pay to protect tropical rain forests. *Land Economics* 73, 196–210.
- Olsen, J.A., 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), 1–12.
- Ready, R.C., Buzby, J.C., Hu, D., 1996. Differences between continuous and discrete contingent value estimates. *Land Economics* 72, 397–411.
- Ryan, M., Scott, D.A., Donaldson, C., 2000. Valuing health care using willingness-to-pay: a comparison of the payment card and dichotomous choice methods. Paper presented at the Sixth Meeting of EuroWill Group, Dublin.
- Smith, R., 2000. The discrete choice willingness to pay question format in health economics: should we adopt environmental guidelines. *Medical Decision Making* 20, 194–206.
- Tversky, A., Sattath, S., Slovic, P., 1988. Contingent weighting in judgment and choice. *Psychological Review* 95, 372–384.