Title	Do Ordering Effects Matter in Willingness-to-pay Studies of Health Care?
Author(s)	Stewart, Jennifer; O'Shea, Eamon; Donaldson, Cam; Shackley, Phil
Publication Date	2000
Publication Information	Stewart, J., O'Shea, E., Donaldson, C., Shackley, P. (2000) Do Ordering Effects Matter in Willingness-to-pay Studies of Health Care? (Working Paper No. 0046) Department of Economics, National University of Ireland, Galway.
Publisher	National University of Ireland, Galway
Item record	http://hdl.handle.net/10379/1193

Do Ordering Effects Matter in Willingness-to-pay Studies of Health Care?

Jennifer M. Stewart[1]   Eamon O'Shea[2]   Cam Donaldson[3]   Phil Shackley[4]

Working Paper No. 46   **October 2000**

_____

[1] Corresponding Address: Jennifer M. Stewart, Department of Economics, National University of Ireland, Galway. e-mail: jennifer.stewart@nuigalway.ie

[2] Department of Economics, National University of Ireland, Galway

[3] University of Calgary

[4] University of Sheffield

Abstract

Willingness-to-pay studies are increasingly being used in the evaluation of health care programmes. There are, however, 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 simultaneously asking willingness-to-pay questions about three health care programmes, does the order the programmes are presented have an impact on the reported willingness-to-pay? We present the results from a survey which allowed us to test for ordering effects and examine, in particular, if the respondent's past experience with the health care service interacted with the ordering effects.

1.   Introduction

Work on eliciting willingness-to-pay (WTP) values in order to evaluate projects has been done extensively in environmental economics and is a growing field in health economics (Diener et al., 1999). In 1993, the National Oceanographic 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 methods 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 assessments of health care programmes as part of a larger research project (EuroWill)[1] to examine WTP methods in the health care field. This paper examines the consequences of varying the order in which WTP questions on three different health care programmes are presented to respondents. The empirical question addressed is whether WTP for the provision of a given good depends upon its position in a sequence of n goods where n > 1. It is assumed that asking people to value a number of different programmes from a sequence will lead to different results than if the various programmes were estimated in isolation. However, whether or not the sequence value is more valid than the isolated value is far from clear from either economic theory or psychological theory and is not addressed in this paper. The focus in this paper is on the interaction between the order of presentation of the three programmes in the sequence and the reported WTP.

In the following section, previous studies addressing this issue are reviewed in both health and environmental economics. Then the conceptual issues to be addressed in this paper are specified. The next section includes a description of the survey design used to address our empirical questions. We then present our data analysis.

---

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

2. Background

Asking people to put a monetary WTP value on a sequence of goods or services will likely yield different results depending on the order of the goods or services because answers to earlier questions will affect responses to later questions (Gorden, 1969). The literature on ordering effects is sparse, however, with only a small number of published studies having examined whether an ordering effect arises when multiple CV questions are asked in a single survey instrument. Brookshire et al. (1981) were the first to report results, albeit from a small sample, indicating that if more than one good is presented in a valuation sequence then the ordering of that sequence could affect the value assigned to any particular good. In contrast a study by Boyle et al. (1990) found that the order of CV questions did not have a significant effect on estimated Hicksian surplus, though, once again, the sample size was small.

In a subsequent study, Boyle et al. (1993) examined the impact of water flows on WTP for white-water boating trips on the Grand Canyon and again considered the possibility of ordering effects. A dichotomous choice question was used. Respondents were asked whether or not they would be willing to pay a specific stated amount for the described trip. The amount was varied between respondents and the researchers could then estimate the demand function. Boyle et al. (1993) found that the order of the questions did have an impact for respondents with less experience with boating trips, while for respondents with more experience no effect was observed.

Kartmann et al. (1996) tested for ordering effects in a health care context, also using a dichotomous choice question. They surveyed patients with reflux oesophagitis for their WTP for three different treatments. They did not find any impact of changing the order in which the treatments were presented. This result is similar to the Boyle et al. (1993) result in that respondents with previous experience with the service were not influenced by the order in which scenarios were presented.

Halvorsen (1996) examined the impact of the order of questions in a slightly different framework than the previous articles. Rather than comparing different programmes/scenarios, Halvorsen was interested in decomposing the total WTP for an

emission reduction programme into benefits from improved health and from reduced environmental damage. Respondents were asked, in an open-ended question, to state the maximum amount they would be willing to pay for the programme because of either health benefits or environmental benefits. Then they were asked their WTP for the remaining category of benefits. The results indicated that the order of the questions did have an impact on the reported WTP for each aspect of the benefits, but that the total WTP was not affected by the order of the questions.

Based on the results from this previous research, we not only test whether the order of the questions had an impact on the full sample, but we also break our sample into groups by previous experience with the health condition. It should also be noted that many of the previous papers have used dichotomous choice questions while we used a payment card. It may be useful in future work to examine if the structure of the WTP question has an impact on the probability of observing ordering effects.

3.   Conceptual Issues

The study used a split sample approach defined by the order of WTP questions for three health care programmes as shown in Table 1. The three health care programmes are cancer (*C*), hearts (*H*) and community care (*CC*). The superscripts in Table 1 denote the order of the good within a sequence, while the subscripts refer to the particular version of the survey. In order to test for ordering effects, respondents were randomly assigned to one of two samples of the survey. In the first sample respondents were asked their WTP for the cancer programme first, then the hearts programme, and, finally, the community care programme. In the second sample of the survey, the order of the programmes was reversed. Respondents were first asked their WTP for the community care programme, then the hearts programme, and, finally the cancer programme. By comparing the reported WTP values in each survey, it was possible to test for ordering effects.

Respondents were given a detailed description of the effects of the health care programmes they were asked to value, summarised as follows: an increase in a pain treatment programme for cancer patients, an increase in heart operations, and an increase

in community care services for dependent elderly patients (The full descriptions of the programmes which were presented to respondents are included in Appendix A). The questions were put in the form of an 'either/or' sequence rather than an 'additional good' sequence. With an 'either/or' sequence respondents were presented with an ordering of health care programmes (*C, H* and *CC* or *CC, H* and *C*) in advance and asked to state their WTP valuation of each. Respondents were informed that the programmes were in competition with each other for funding. Respondents were explicitly asked to ignore the other two programmes when providing a WTP estimate for any one of the programmes.

Under rational choice theory, preferences should not depend upon the order in which options are presented to the individual in any sequence. An ordering effect would, therefore, violate the principle of procedural invariance of rational choice theory (Kahneman and Tversky, 1984; Tversky et al., 1988). A full account of theoretical expectations regarding ordering effects and context independence, particularly for nested goods, is given by Carson and Mitchell (1995) in their discussion of sequencing and nesting issues in CV studies. For 'additional goods' sequences, the economic argument relies primarily upon the income and substitution effects which may occur when a list of purchase possibilities is expanded (Carson and Mitchell, 1995; Carson et al., 1998).

Carson and Mitchell (1995) also note that extending the sequence also increases the cognitive effort demanded of respondents, resulting in a tendency from study designers to reduce the detail provided for programmes later in the sequence. This tendency may inadvertently lead to respondents perceiving such briefly described goods as less credible resulting in lower stated values for goods within sequences than within single good studies. The main psychological based interpretation of sequencing behaviour is based on moral satisfaction (Kahneman and Knetsch, 1992) or warm glow (Andreoni, 1990) motives. The first good or service in a sequence will likely have captured most of the moral satisfaction or warm glow effects leading to the observed ordering effects.

In contrast to 'additional goods' sequences, economic theory provides no expectation of ordering effects within 'either/or' sequences. The absence of theory means that when and where these effects occur they must relate either to the psychological based interpretation or, alternatively, respondents perceive the various programmes as 'additional' rather than

different question and the whole basis of the WTP exercise is problematic. If the psychological interpretation is the problem then a high WTP for the first programme will be associated with a relatively high WTP for a subse-quent programme. Conversely, a low WTP for the first programme will be associated with a relatively low WTP for a subsequent programme.

In this paper we considered four hypotheses as to why the order of the programmes may have an impact on the reported WTP values and a fifth hypothesis related to the respondents' previous experience.

H_1: Starting Point Bias: the stated WTP for the second and third programmes in the sequence are positively related to the level of stated WTP for the first good presented in the sequence. A higher initial value leads to an upward bias for subsequent values and a lower initial value leads to a downward bias for subsequent values.

Under this hypothesis, the survey, which begins with the higher ranked programme, should elicit higher WTP values than the survey which begins with the lower ranked programme. If the ranking of programmes is such that the cancer programme is preferred to the community care programme, regardless of the order of the programmes in the sequence, then a starting point effect occurs if the WTP valuation yields the following result: $CC_1^3 > CC_2^1, C_1^1 > C_2^3$, and $H_1^2 > H_2^2$. Conversely, if the ranking of programmes is such that the community care programme is preferred to the cancer programme then a starting point effect occurs if the WTP valuation yields the following results: $CC_2^1 > CC_1^3, C_2^3 > C_1^1$, and $H_2^2 > H_1^2$.

H_2: Binding Budget Constraints: budget constraints become more binding as respondents move through a sequence.

As the survey progresses, respondents are likely to be aware that they have less to spend after they have already reported contributions for other programmes. Therefore, the reported WTP for a programme may be lower if it is last in the sequence than if it is first in the sequence. Strictly speaking, this effect should not happen in this survey, as respondents are asked to treat the three programmes as substitutes for each other. The

expectation is, therefore, that respondents will not 'add up' their WTP values, but, rather, will treat them independently. However, there is no guarantee that people will act in this way, particularly because respondents to the survey are also reminded in making the valuation to consider that their contribution will reduce what they have left to spend on other things. Respondents may treat the sequence as 'additional' rather than 'either/or'. If budget constraints are binding we would expect the following results: $C_1^1 > C_2^3, CC_2^1 > CC_1^3,$ and $H_1^2 = H_2^2.$ We would expect that respondents with higher incomes would be less conscious of a budget constraint and would, therefore, be less influenced by the order of the programmes. However, we would not expect the starting point bias to affect respondents with higher incomes any less than it would respondents with lower incomes. 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 and can test this hypothesis by examining the interaction between survey type and income.

H₃: Moral Satisfaction or Warm Glow: the first programme in the sequence is likely to capture the majority of moral satisfaction or warm glow values resulting in the observed sequencing effects.

Ordering effects due to moral satisfaction or warm glow, as predicted by psychological theory, would yield a similar result to hypothesis H_2, i.e. $C_1^1 > C_2^3, CC_2^1 > CC_1^3,$ and $H_1^2 = H_2^2.$ Warm glow effects are likely to be mostly captured by the first programme in a sequence and diminish as more programmes are added to the valuation process. The first programme in the sequence is the most important programme for respondents because it provides the initial opportunity for them to demonstrate their willingness to help, thereby exhausting the bulk of the moral satisfaction associated with the decision to support a good cause. The health care programme is valued for the moral satisfaction associated with the contribution rather than as an end in itself.

H₄: Yea-saying/Nay-saying: respondents are more likely to say 'yes' once they have already said 'yes' and more likely to say 'no' when they have already said 'no'.

The fourth hypothesis is that there may be a tendency towards 'yea-saying/nay-saying'. If first asked to value a programme that they do not want, respondents may be inclined,

subsequently, to respond in a similar negative way to programmes that they actually like or want. Conversely, beginning the WTP valuation process with a positive WTP value may lead to a positive value for all programmes, even programmes which respondents do not like or want. The order of the programmes may, therefore, have an impact on the probability of reporting a positive WTP. This hypothesis can be tested in a straightforward manner by examining whether proportions of zero WTP responses differ between sequences. We predict that the survey which begins with the higher ranked programme, will result in a lower proportion of zero WTP values for the other programmes than the survey which begins with a lower ranked programme.

H_5: *Previous Experience: respondents with previous experience with the good have better knowledge of the value of the good to themselves and will not be influenced by the order of the WTP questions.*

Previous research indicated that respondents with previous experience of the good were not influenced by the order of the WTP questions. These respondents may have better knowledge of the value of the good to themselves and are not influenced by the order of the questions. We divide our sample into groups by whether or not they have had previous experience with the health condition. These groups are not mutually exclusive in that a respondent may have had previous experience of more than one health care programme or condition.

Finally, there is the issue of the internal consistency of the results. More specifically, are the elicited WTP values correlated with the socio-economic, health and demographic details of the respondents as predicted by theory? In forming our expectations we need to consider why respondents are willing to contribute to a health care programme. People may be willing to contribute for selfish reasons, that is, respondents have the expectation that they, or their family, may benefit from the programme. They, or their family, may have benefited from the programme in the past. There could be altruistic reasons, that is, respondents are interested in contributing to a social good. Furthermore, respondents may be willing to contribute because they derive utility from just knowing that the programme exists.

We predict that income is likely to be positively correlated with WTP. This prediction is based on the belief that respondents consider their budget constraint when reporting their WTP (see H2). There may be two possible exceptions. First, behaviour, which leads to diseases, may vary by social class, and income may proxy for social class, thus the correlation between WTP and income could be due to different expected needs in health care programmes. Second, respondents with higher incomes are more able to purchase 'protective' lifestyles and, therefore, be at lower risk of needing the provision of health care. In both of these cases we may observe a negative correlation between income and WTP. Despite the last two possibilities, overall, we predict a positive correlation.

The sex of the respondent may be correlated with their WTP. For example, in the description of the community care programme it is explicitly stated that, generally, elderly women would receive the extra service, while, for the heart programme, it was stated that the average recipient of an operation would be male. If respondents are motivated by selfish reasons, that is, they are willing to contribute because they expect to benefit from the programme, then women should be willing to contribute more to the community care programme than men and vice versa for the hearts programme.

Age may be correlated with the WTP if older respondents have a higher expectation of using the programme. We also expect that the relationship is nonlinear and decreasing with age, that is, the difference in WTP between a twenty-year-old and a forty-year-old is greater than between a sixty-year-old and an eighty-year-old. Marital status may be correlated with the WTP. Respondents, who are single, may have different attitudes toward health care programmes than respondents who are married. However, a priori, we cannot make predictions about the direction of the impact on WTP. Education may also be correlated with the WTP. Respondents with different education levels may have different attitudes toward health care programmes. Again, as with marital status, a priori we cannot make predictions about the direction of the impact on WTP.

The respondent's perceived risk and respondent's past experience would be correlated with WTP if people's willingness to contribute is dominated by selfish reasons. We expect that respondents with a higher perceived risk and past experience of a programme would have a higher WTP for that programme. Health status will be correlated with WTP

We would expect a negative correlation between high health status and WTP. Being a smoker will be positively correlated with WTP if respondents who smoke have a higher expectation of using the services than those who do not smoke.

4. Survey Method

The survey was carried out in the Western Health Board region of Ireland, which contains a population of approximately 350,000 people and includes the counties of Galway, Mayo and Roscommon. Both the pilot survey and the main survey were carried out by the Economic and Social Research Institute (ESRI), which is the primary research agency for economic and social research in Ireland. The sample design was based on a two-stage clustered sample using the Electoral Register as a population frame. At the first stage of the sample collection, the population across District Electoral Divisions (DED's) is aggregated up to a prespecified minimum cluster-size threshold. These initial clusters, or primary selection units, form the first stage of the sample selection. The second stage of the sample selection relates to the selection of individuals (electors) from within each of the primary selection units. Only the named individuals from each of these units can be interviewed, though the questions asked relate to their household.

To ensure representativeness of the data, it was necessary to adjust the composition of the effective sample in such a way as to eliminate any identifiable bias, which may have arisen from sample design effects. Sample surveys can also be affected by differential non-response among various sub-groups of the population. In particular, response rates may be lower among 'lower' socio-economic groups and groups with less interest in the issue at hand (Bateman and Langford, 1997). In this survey greater representativeness is achieved by re-weighting the electoral data so as to ensure that the structure of the sample corresponds with the known structure of the population of the Western Health Board, as derived from the most recent population Census. The data generated from the electoral register is re-weighted on the basis of the principal economic status of head of household, household composition and sub-regional classification. The survey remains a cluster sample but the over-all population frame from which the cluster is drawn is now representative of the overall population in the region.

A total of 473 people were approached for interview. The response rate was 45 per cent giving a total number of people interviewed of 214. The main difficulty for the interviewers was meeting the people selected for interview face-to-face. Once contact was made the response rate was high with only 8 per cent of people refusing to be interviewed when met face-to-face by the interviewers. All respondents were interviewed in their own homes. The main problem was that almost 50 per cent of the sample could not be contacted, or located, or were temporarily absent when the interviewers called. Interviewers did carry out one return visit when named respondents were absent from the households but this did not succeed in raising the response rate significantly. One of the problems with using the electoral register as the population frame is that many younger people in Ireland return home to vote but live elsewhere in the country. This is particularly true of the Western region, which has seen very high levels of outward migration in recent decades.

In terms of survey design, respondents were first asked about their perception of ever needing the programmes and their past experience with the health states. Next, respondents were asked to consider the relative importance of each programme and then to rank the programmes from most important to least important. The next section of the survey asked the respondents about their WTP for each programme. This section began by asking respondents if they would be willing to contribute anything in extra taxation for the given expansion in the programme. 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, in the absence of private markets. If the respondent answers no to both of these questions, they are asked to explain the reasons why they are 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 card included in Appendix A of this paper.

5.   Data Analysis

Table 2 lists the variables used in our analysis and provides a description of each variable.

Table 3 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 and also ranked least important by a smaller proportion of respondents. We concluded that the cancer programme was the most important programme according to respondents. The heart programme was ranked most important by a smaller proportion of respondents than ranked community care most important, but it was ranked second most important by a larger proportion of respondents than ranked community care second most important. The heart programme was ranked least important by a smaller proportion of respondents than ranked community care least important. It was difficult to decide which of these two programmes was ranked higher than the other. We concluded that the heart programme was ranked higher than community care and that this ranking was consistent across both samples.

Table 3 indicates that there was a difference in the proportions in each of the rankings between the samples. When we performed an ANOVA analysis, we found that there was a statistically significant difference between the samples in the proportions that ranked the cancer programme most important and for the proportions that ranked the cancer programme least important. 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. A lower proportion ranked the cancer programme least important in the first sample. This result provides clear proof of an ordering effect.

None of the other differences in the proportions were statistically significant between the samples.

At the point in the survey where respondents were asked to rank the programmes, they had not yet been asked any WTP questions. However, they had been asked 'warm up' questions pertaining to their own perception of their own risk for the three conditions and to their previous experience with the three conditions. The order of these risk and experience questions was switched between the two samples. The ranking questions were asked following the 'warm up' questions. In the first sample, respondents were first given the description of the cancer programme, then the description of the heart programme and then asked to rank the two programmes. Next, they were given the description of the community care programme and then asked to rank all three programmes. In the second sample, the positions of the cancer and community care programmes were switched. First, respondents were given the description of the community care programme, then the description of the heart programme, and then asked to rank the two programmes. Next, they were given the description of the cancer programme and asked to rank all three programmes. At this early point in survey we observed ordering effects and in a relatively simple exercise compared to reporting WTP.

Table 4 presents the unconditional means of the WTP for the three programmes for each sample. It is interesting to note that the implied ranking of the programme by the mean WTP changes between the two samples. The implied ranking in the first sample was cancer, hearts, then community care and in the second sample, the implied ranking was community care, cancer, then hearts. This change in the rankings is suggestive of an ordering effect. The mean WTP for the cancer programme was lower in the second sample and the mean WTP for the community care programme was higher in the second sample: $C_1^1 > C_2^3$ and $CC_2^1 > CC_1^3$. This result is consistent with the predictions from our second and third hypotheses of increasing budget constraint awareness and warm glow effect, respectively. Our first hypothesis of a starting point bias is not supported by this result.

To avoid bias from distribution assumptions, we calculated a non-parametric 95% confidence interval using the bootstrap method (Efron and Tibshirani, 1993) using 1000

simulations. We report the bias-corrected confidence interval. As can be seen from Table 4, the median was less than the mean suggesting that the sample was positively skewed. The bias-corrected boot-strap method corrects the confidence interval for this skewness. Using the bootstrap confidence interval, 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. $\hat{H}_0 : CC_1^1 = C_2^3, H_1^2 = H_2^2$, and $CC_1^3 = CC_2^1$. Similarly, the null hypothesis that the mean WTP of the three programmes were equal, i.e. $\tilde{H}_0 : C_1^1 = H_1^2 = CC_1^3$ and $C_2^3 = H_2^2 = CC_2^1$, could not be rejected for both samples.

We have some concerns about the nature of our data that would not allow us to use a typical OLS regression. Our first concern about our data was that it was censored. Respondents were presented with a payment card that had a maximum value of £200 and included a space to allow respondents to enter an amount higher than £200. However, none of the respondents filled in a higher amount although several indicated a WTP of £200. It was also the case that the data was censored at zero. Respondents were unable to indicate a negative WTP.

Our second concern about the data was that not only was it censored, but that it was discrete in nature rather than continuous.[2] All the respondents who indicated a positive WTP stated an amount that was on the payment card. We take these responses (a_j) to indicate that the respondent was willing to pay the amount they indicated but not the next highest amount (a_{j+1}) on the payment card. We assume that the respondent's true WTP (y_i^*) lies somewhere in this interval, $a_j \leq y_i^* < a_{j+1}$.. The observed WTP (y_i) is related to the true WTP by the following set of relationships:

$$
\begin{aligned}
y_i &= a_1 & \text{if} & & y_i^* < a_2 \\
y_i &= a_2 & \text{if} & & a_2 \leq y_i^* < a_3 \\
&\vdots & & & \vdots \\
y_i &= a_{n-1} & \text{if} & & a_{N-1} \leq y_i^* < a_N \\
y_i &= a_N & \text{if} & & a_N \leq y_i^*
\end{aligned}
\tag{1}
$$

We assume that the true WTP is a function of observable characteristics, $y_i^* = x_i'\beta + u_i..$
Then the probability of observing $y_i = a_j$ is:

$$
\begin{aligned}
\Pr[y_i = a_j] \quad &= \quad \Pr[a_j \leq y_i^* < a_{j+1}] \\
&= \quad \Pr[a_j \leq x_i'\beta + u_i < a_{j+1}] \quad\quad\quad (2) \\
&= \quad \Pr[a_j - x_i'\beta \leq u_i < a_{j+1} - x_i'\beta]
\end{aligned}
$$

If we assume that the errors follow a normal distribution with a mean of zero and variance of σ^2, then we get:

$$
\Pr[y_i = a_j] = \Phi[a_{j+1} - x_i'\beta/\sigma] - \Phi[a_j - x_i'\beta/\sigma] \quad\quad\quad (3)
$$

where Φ denotes the cumulative normal distribution. The log-likelihood function is then:

$$
\ln L = \sum\sum \ln(\Pr[y_i = a_j]). \quad\quad\quad (4)
$$

One final issue in our empirical methodology was how to interpret a zero WTP. An important proportion of the respondents indicate that they are unwilling to contribute to the programme. Previous research (Olsen and Donaldson,1998) has indicated that the reported WTP of zero does not necessarily indicate a true WTP of zero, but rather some respondents 'protest' against paying for the programmes. To determine why respondents were unwilling to contribute, we asked them to explain why they would not be willing to pay for the programme. Two of the possible answers were "This programme is of no value to my household" and "I can't afford it". If these were the reasons the respondent reported a zero WTP then we concluded that zero was 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 also on the full sample and the results do not change.

Table 5 contains the results from the interval regressions[3] for each of the three programmes. The coefficient signs were generally of the sign we predicted although most were not statistically significant. Women were less willing to pay for the heart and cancer programmes and more willing to pay for the community care programme although the coefficient was not statistically significant. Age had a positive, but decreasing, impact on the WTP for all programmes, but, again, the coefficient was not statistically significant. Smokers had a higher WTP, singles had a lower WTP, and respondents with less that good health status had a higher WTP for all three programmes but the coefficients were not statistically significant. Respondents with a primary certificate had a lower WTP for all programmes and the coefficient was statistically significant for the cancer and heart programmes although not for the community care programme. Income had a positive significant impact on WTP for all three programmes as we had predicted.

To examine ordering effects we look to the coefficients of the sample variable. These coefficients indicate the impact of question order on the WTP for each programme after controlling for observed potential differences between the samples. For the cancer and hearts programmes the coefficients were negative indicating that the WTP for these programmes was lower in the second sample. For the community care programme the coefficient was positive indicating that the WTP was higher in the second sample. These coefficients were not statistically significant, but it is interesting to take a look at how the results correspond to our hypotheses. The observed relationships are: $C_1^1 > C_2^3$ and $CC_2^1 > CC_1^3$. This pattern does not support our first hypothesis of a starting point bias because, if the first hypothesis was true, we would predict that $CC_1^3 > CC_2^1$. The observed relationships do support our second and third hypothesis. To further distinguish between the second and third hypotheses, we also ran regressions that included an interaction term between the sample variable and income. We predict that if the budget constraint hypothesis is true then we would observe that order effects do not affect respondents with higher incomes. We do not present the results here from these regressions, because the coefficients on other variables were the same as in Table 5 and the interaction term was never statistically significant. This result does not support the second hypothesis and we

---

[3] We also ran linear OLS, ordered probit, and ordered logit models. The results are similar for all models.

conclude that the warm glow hypothesis is the best explanation for our results. This result also indicates that respondents were viewing the WTP questions as 'either/or' questions.

With this type of model it is difficult to discuss how well it fits the data. We have reported the log-likelihood, the null log-likelihood, and the likelihood ratio index, which is a ratio of the two log-likelihoods. Our model did explain some of the observed variation and it did this the best for the community care programme. In addition, the signs of the coefficients are consistent with standard economic theory predictions.

To examine further the possible effect of question order we examined the impact on the probability of reporting a zero WTP. Table 6 shows the proportion of the respondents that reported a zero WTP for each programme by sample. It was apparent that fewer of the respondents reported a true zero in the second sample. We again performed an ANOVA analysis, this time to test if the proportion of reported zeros was different between the two samples. The results indicated that the difference for the heart and community care programme was statistically significant, but for the cancer programme the difference was not statistically significant.

Our first hypothesis of a starting point bias would predict that, when the community care programme was presented first, the reported WTP values would be pulled downwards because it was preferred less to the other programmes. We expect more zero values in the second sample if this hypothesis is correct. We did not observe this result, so we concluded that the ordering effects did not occur because of a starting point bias. Our second hypothesis of an increasing awareness of a budget constraint and third hypothesis of warm glow would predict that when a programme was presented last that there would be more zero values. Our result that more zero values were reported for community care in the first sample was consistent with this prediction. However, the result for the cancer programme was not consistent with this prediction. Our fourth hypothesis of a yea-saying/nay-saying response would predict that there would be more zeros for all programmes in the second sample because the community care programme was the lower ranked programme. Our results do not support our fourth hypothesis either. We do not have an explanation for why there would be fewer zero values reported when the least preferred programme was presented first.

Our fifth hypothesis was that the order in which the programmes were presented would not have an impact on those respondents with previous experience. A reason for this expectation is that respondents with previous experience may have a better idea of the value of the programme to them and, therefore, may be more resolute in their valuation of the programme. Table 7 presents the coefficients for the sample variable from regressions on the WTP for each programme separating the respondents by their previous experience with the programmes. Our prediction was that the coefficient for the sample variable would not be statistically significant for the respondents with previous experience.

For the hearts and community care programmes, the coefficient was not statistically significant for respondents who had previous experience with the relevant programme, which was as we predicted. However, we found an unexpected result for the cancer programme. Respondents with previous experience with cancer had a statistically significant lower WTP for the cancer programme when it was presented last. This result contradicted our predictions regarding previous experience.

One explanation for this unexpected result may be related to the fact that the cancer programme is an invasive procedure. When we present the community care programme first, respondents, who have experience with cancer, may have realized that they would prefer a less invasive approach, such as a community care-based programme of treatment. To determine if there was an interaction between reported WTP and the other programmes presented, we also examined the WTP for the heart and community care programmes by whether or not the respondents had previous experience with cancer. Our prediction was that if this explanation was correct then the WTP for community care of the respondents with previous experience with cancer would be higher in the second sample. However, the coefficient on the sample variable in this regression was negative and not statistically significant.

Interestingly, those with previous experience of community care had a statistically significant lower WTP for both the cancer and heart programme when the community care programme was presented first. This result was again suggestive of the possibility that there was an interaction between the programmes.

When we examine the WTP for each programme by whether or not respondents had previous experience with a heart problem, we did not find that the question order had a statistically significant impact.

6.   Conclusion

Did we observe an impact of the question order on the reported WTP? There is evidence of an ordering effect. Respondents in the first sample (*C, H, CC*) were significantly more likely to rank the cancer programme as their first choice than were respondents in the second sample (*CC, H, C*). Moreover, the mean WTP for the cancer programme was lower in the second sample and the mean WTP for the community care programme was higher in the second sample. However, using the bootstrap confidence interval, we could not reject the null hypothesis that the mean WTP for a programme in the first sample was equal to the mean WTP in second sample, for all programmes.

We proposed five hypotheses that might explain the observed order effect. The first hypothesis of a starting point bias was not supported by the observed pattern of the reported WTP values. The second hypothesis of an increasing budget constraint awareness and the third hypothesis of a warm glow effect were supported by the pattern of the WTP values. When we included an interaction term between the sample and income, it did not have impact on the WTP values. Therefore, we concluded that the observed pattern of WTP values was best explained by a warm glow effect. The warm glow effect might occur because health care is a good that people are familiar with contributing towards (and being asked to contribute towards). Therefore, even though they are presented with three competing programmes, the respondents feel that they have met their social obligations once they contribute to one programme.

We observed that there were fewer zero values reported when the community care programme was presented first. This result does not correspond to any of our hypotheses and indicates that the effect of question order is more complex than our simple hypotheses can explain.

Previous research found evidence that the question order impact was not present for respondents with previous experience. We found that there is some interaction between previous experience and the order of the programmes, but not in the manner we expected. For the cancer programme, respondents with previous experience report a lower WTP for the programme when it is presented last, while respondents with no previous experience were not influenced by the programme order. This result raises questions about whom to survey for WTP estimates and the possible existence of interactions between the programmes that are being analysed in the health care field.

References

Andreoni, J. 1990. Impure altruism and donations to public goods: a theory of warm-glow giving. *Economic Journal*, 100: 464-477.

Arrow, K., R. Solow, P.R., Portney, E.E., Leamer, R., Radner and H. Schuman. 1992. Report of the National Oceanographic and Atmospheric Administration Panel of Contingent Valuation. *Federal Register*, 58:4601-14, Washington D.C.

Bateman, I.J. and I.H. Langford. 1997. Budget constraint, temporal and ordering effects in contingent valuation studies. *Environment and Planning*, 29: 1215-1228.

Boyle, K.J., S.D. Reiling and M.L. Phillips. 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., M.P. Welsh, and R.C. Bishop. 1993. The Role of Question Order and Respondent Experience in Contingent-Valuation Studies. *Journal of Environmental Economics and Management*, 25 (S-80- S-90).

Brookshire, D.S., R.C. d'Arge, W.D. Schulze and M.A. Thayer. 1981. Experiments in valuing public goods. In V.K. Smith, (ed.). *Advances in Applied Microeconomics: Vol 1*, Greenwich, CT: JAI Press.

Carson, R.T. and R.C. Mitchell. 1995. Sequencing and nesting in contingent valuation studies. *Journal of Environmental Economics and Management* 28: 155-173.

Carson, R.T., N.E. Flores, and W.M. Hanemann, W.M. 1998. Sequencing and valuing public goods. *Journal of Environmental Economics and Management*, 36:314-323.

Diener, A., B. O'Brien and A. Gafni, 1996. Contingent Valuation Studies in the Health Care Literature: A Review and Classification. Working Paper, Centre for Health and Policy Analysis, McMaster University, 1996.

Diener, A.A., C. Donaldson, B.J. O'Brien and A. Gafni. 1999. *Contingent Valuation in Health Care: A Caution Against Uncritical Adoption of the NOAA Guidelines*. Manuscript.

Donaldson, C., A.M. Jones, T.J. Mapp, and J.A. Olsen. 1998. Limited Dependent Variables in Willingness-to-pay Studies: Applications in Health Care. *Applied Economics*, 30: 667-677.

Donaldson, C., P. Shackley, and M. Abdalla. 1997. Using Willingness-to-pay to Value Close Substitutes: Carrier Screening for Cystic Fybrosis Revisited. *Health Economics*, 6:1145-159.

Donaldson, C., R. Thomas, D.J. Torgerson. 1997. Validity of open-ended and payment scale approaches to eliciting willingness-to-pay. *Applied Economics*, 29: 79-84.

Donaldson, C., P. Shackley, M. Abdalla, and Z. Miedzybrodzka, 1995. Willingness-to-pay for Antenatal Carrier Screening for Cystic Fybrosis. *Health Economics*, 4: 439-452.

Efron, B. and R.J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Boca Raton, FI: Chapman and Hall.

Farrar, S. and M. Ryan. 1999. Response-Ordering Effects: A Methodological issue in Conjoint Analysis. *Health Economics*, 8: 75-79.

Gorden, R. 1969. *Interviewing: Strategy, Technique and Choices*. Homewood, Il: Dorsey.

Halvorsen, B. 1996. Ordering Effects in Contingent Valuation Surveys: Willingness-to-pay for Reduced Health Damage from Air Pollution. *Environmental and Resource Economics*, 8: 485-499.

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

Kahneman, D. and A. Tversky. 1984. Choices, values and frames. *American Psychology*, 39: 341-350.

Kartman, B., N-O. Stalhammar and M. Johannesson. 1996. Valuation of Health Changes with the Contingent Valuation Method: A Test of Scope and Question Order Effects. *Health Economics*, 5: 531-541.

Olsen, J. A. and C. Donaldson, 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., J. Pereira, C. Donaldson, K. Kidholm, and P. Shackley. 1999. *The Insensitivity of the Size of the Good for 'Willingness-To-Pay': New Evidence for Health Care*. Manuscript prepared for EuroWill meeting.

Ryan, M., D.A. Scott, and C. Donaldson. 1999. *Valuing Health Care Using Willingness-to-pay: A Comparison of the Payment Card and Dichotomous Choice Methods*. Manuscript prepared for EuroWill meeting.

Scott, A. and Sandra V. 1999. Patients, Doctors and Contracts: An Application of Principal-Agent Theory to the Doctor-Patient Relationship. *Scottish Journal of Political Economy*, 46:111-134.

Tversky, A. and D. Kahneman, 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185: 1124-1131.

Tversky, A., S. Sattath and P. Slovic, 1988. Contingent weighting in judgement and choice, *Psychological Review*, 95: 372-384.

Appendix A - Description of Programmes

PAIN TREATMENT FOR CANCER PATIENTS

200 more patients with advanced cancer could have pain 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 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 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 of 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 the people benefiting from this programme will be women aged 75 years and over.

PAYMENT CARD

Amount		Amount		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	
£20.00		£100.00		(Please specify)	

In the interview please tick (p) the amounts you are sure you would pay.

In the interview please put a cross (X) beside the amounts you are sure you would not pay.

In the interview please put a circle (O) around the amount which is the maximum you would be willing to pay.

Table 1: Split Sample Design

Sample   Order of WTP Questions
1   Cancer (C_1^1) Heart (H_1^2), Community Care (CC_1^3)
2   Community Care (CC_2^1), Heart (H_2^2), Cancer (C_2^3)

Table 2: 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	=	income 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). The income was taken as the log of the midpoint of the income interval.
Survey 2	=	1 for survey 2 where the order of the programmes was community care, hearts, and cancer, 0 for survey 1 where the order of the programmes was cancer, hearts, and community care.
Experience	=	1 if answered yes to "Have you or anyone in your close family ever had personal, first hand experience of (the relevenat condition)?"

Table 3: Ranking of Programmes by Survey Type[a] in Percentage of Sample (number of respondents in parentheses)

	Cancer		Hearts		Community Care	
	1	2	1	2	1	2
Most Important	79.65 (n=90)	63.67 (n=64)	47.79 (n=54)	50.50 (n=51)	52.21 (n=59)	52.48 (n=53)
Second Most Important	17.79 (n=20)	26.73 (n=27)	35.40 (n=40)	31.68 (n=32)	11.50 (n=13)	14.85 (n=15)
Least Important	2.65 (n=3)	9.90 (n=10)	16.81 (n=19)	17.82 (n=18)	36.28 (n=41)	32.67 (n=33)

[a] Order in Survey 1; Cancer, Hearts, Community Care. Order in Survey 2; Community Care, Hearts, Cancer

Table 4: Descriptive Statistics of Willingness-to-Pay by Survey Type[a]

	Cancer	**Hearts**	**Community Care**
Survey 1			
Mean	49.22	42.37	37.14
Std. Dev.	61.56	56.36	54.54
Median	20.00	20.00	10.00
BS 95% C.I.	38.08–62.65	32.34–54.64	27.34–48.88
Zeros	11	15	13
n	102	99	96
Survey 2			
Mean WTP	41.40	37.13	42.12
Std. Dev.	51.81	48.04	51.41
Median	20.00	20.00	20.00
BS 95% C.I.	31.26–52.22	28.15–49.29	32.05–53.95
Zeros	5	4	2
n	86	80	91

[a] Order in Survey 1; Cancer, Hearts, Community Care. Order in Survey 2; Community Care, Hearts, Cancer.

Table 5: Interval Regression of Willingness-to-pay for Programmes
(standard errors in parentheses)

	Cancer	Heart	Community Care
Female	-6.68	-10.86	7.67
	(10.23)	(10.26)	(9.26)
Age	1.70	1.15	0.88
	(1.93)	(1.97)	(1.73)
Age-squared	-0.02	-0.02	-0.01
	(0.02)	(0.02)	(0.02)
Single	-22.10	-17.70	-11.55
	(14.18)	(13.51)	(12.49)
Primary	-24.99*	-28.61**	-10.03
Education	(14.14)	(13.93)	(12.88)
Own Health	14.78	19.56	13.21
< Good	(13.52)	(13.22)	(12.20)
Smoker	12.47	10.80	5.62
	(12.90)	(12.67)	(11.86)
Income	32.13**	32.15**	32.32**
	(10.34)	(10.86)	(9.39)
Survey 2	-5.37	-5.59	6.96
	(10.05)	(9.92)	(9.02)
Constant	63.57**	60.93**	37.64**
	(11.28)	(11.46)	(10.29)
Sigma	59.86**	57.85**	53.34**
	(3.88)	(3.83)	(3.41)
Log-Likelihood	-504.46	-478.64	-504.30
Null Log-Likelihood	-626.79	-588.13	-633.86
Likelihood Ratio Index	0.195	0.186	0.204
Sample Size	158	151	154
Number of Zeros	16	19	15

* Significant at a 90 % confidence level
** Significant at a 95 % confidence level
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 survey 1.

Table 6: Proportion of Sample Reporting a Zero Willingness-to-Pay(number of zeros in parenthesis)

Cancer		Heart		Community Care	
1	2	1	2	1	2
9.73	4.95	13.27	3.96	11.50	1.98
(11)	(5)	(15)	(4)	(13)	(2)

Table 7: Summary of Coefficient for Survey 2 for Interval Regressions of Willingness to Pay for Programme by Experience with Condition(standard errors in parentheses)

	Cancer	Hearts	Community Care
Experience with Cancer	-32.16*	-24.35	-7.73
	(19.06)	(18.32)	(17.00)
No Experience with Cancer	12.20	2.08	16.24**
	(8.75)	(9.45)	(7.82)
Experience with Hearts	-14.00	-8.77	5.64
	(18.19)	(17.08)	(14.47)
No Experience with Hearts	0.51	0.79	10.69
	(11.03)	(11.15)	(11.10)
Experience with Community Care	-76.25**	-62.29**	-23.47
	(23.51)	(21.18)	(20.22)
No Experience with Community Care	3.74	5.36	13.80
	(10.94)	(11.19)	(10.45)

* Significant at a 90 % confidence level
** Significant at a 95 % confidence level
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 survey 1.