Testing for consistency in willingness to pay experiments

Mandy Ryan *, Fernando San Miguel

Health Economics Research Unit, Department of Public Health, University Medical Buildings, Foresterhill, Aberdeen AB25 2ZD, UK

Received 27 November 1998; received in revised form 20 June 1999; accepted 13 November 1999

Abstract

Given the increased use of willingness to pay (WTP) experiments to elicit the economic value of goods, it is becoming increasingly important to assess the validity of the instrument. Given the problems of testing for external validity in the absence of market, other methods must be developed to test the validity of WTP responses. This paper develops a simple test of consistency in WTP experiments which is based on the theoretical basis of the technique – if commodity A is preferred to B, then individuals should be willing to pay more for A than B. The test is applied to elicit women’s preferences for two alternative treatments for menorrhagia: conservative treatment versus hysterectomy. Thirty percent of respondents failed the consistency test. Cost-based responses were found to partly explain inconsistent responses. This simple test highlights potential problems when using WTP experiments within a cost-benefit analysis framework. Possible solutions to avoid ‘cost-based’ WTP responses are suggested. © 2000 Elsevier Science B.V. All rights reserved.

PsycINFO classification: 3920

JEL classification: C21; C42; C52; D12

Keywords: Willingness to pay; Consistency

* Corresponding author. Tel.: +44-1224-681818 ext 54965; fax: +44-1224-662994.
E-mail address: m.ryan@abdn.ac.uk (M. Ryan).
1. Introduction

Willingness to pay (WTP) experiments have been widely used in environmental economics, and are being increasingly used in health economics, as a method for valuing the benefits of non-marketed commodities (Cummings, Brookshire & Schulze, 1986; Mitchell & Carson, 1989; Carson, Wright, Alberini, Carson & Flores 1994; Diener, O'Brien & Gafni, 1998). Using a survey design, individuals are presented with a hypothetical scenario and asked, in this scenario, for their maximum WTP for the good concerned. ¹

Given that WTP experiments are based on responses to hypothetical questions, tests of their validity must be carried out. Ideally such tests would involve testing the external validity of responses, i.e. the extent to which individuals behave as they say they would. However, in some areas, such as the UK health care system, such tests are difficult to carry out since individuals do not pay at the point of consumption. There is no market in which they can reveal their true maximum WTP. Studies applying WTP experiments to health care have tested for theoretical validity (the extent to which results are consistent with economic theory, or, sometimes, a priori expectation more generally) and convergent validity (the extent to which values elicited using the WTP instrument are similar to values estimated using other techniques) (Johannesson, Jonsson & Borgquist, 1991; Johannesson, Johansson, Kristrom & Gerdtham, 1993; O'Brien & Viramontes, 1994; Donaldson, Shackley, Abdalla & Miedzybrodzka, 1995; Donaldson, Shackley & Abdalla, 1997; Ryan, Ratcliffe & Tucker, 1997).

This paper adds to the literature on the applicability of WTP experiments by looking at the consistency of responses. Given that WTP is based on the value of a commodity, individuals should be willing to pay more for preferred goods. Thus, one way of measuring the consistency of WTP experiments is to ask respondents their preference for the commodities being valued and then compare such preferences with their stated WTP. A respondent that gives a smaller WTP value for their preferred option or greater WTP value for their less preferred option may be considered inconsistent. This test is applied to eliciting women’s preferences for two alternative treatments for menorrhagia:

¹ Willingness to pay experiments are also referred to in the literature as contingent valuation studies.
conservative treatment versus hysterectomy. Following this, possible reasons for inconsistent responses are explored, as are the possible implications of including inconsistent responses to estimate welfare changes. Finally, possible methods to overcome cost-based WTP responses are suggested.

2. Methods

2.1. Subjects and settings

Following a pilot study the main questionnaire was mailed to a sample of 146 women in the Department of Public Health and the Royal Maternity Hospital in Aberdeen. Due to time limitations, no reminders were sent.

2.2. Willingness to pay question

Women were provided with information on menorrhagia and two alternative treatments – hysterectomy and conservative treatment (Pinion et al., 1994; Cameron et al., 1996). Respondents were reminded that any money they spent on menorrhagia treatment would not be available to spend on other goods or services. Following this, the payment card (PC) WTP technique was used to elicit maximum WTP in an experimental situation. Respondents were presented with a range of monetary values and asked to circle the maximum amount they would be willing to pay for each treatment. Based on the results from the pilot study, the PC ranged from £0 to £5000. If individuals answered £5000 they were asked to state the maximum amount they would be willing to pay. They were also asked to give reasons for their answers.

Information was also collected on time to complete the questionnaire, ease of completion (with possible responses on a 1–5 scale, where 1 was very easy and 5 very difficult) and socio-economic characteristics of respondents.

---

2 Menorrhagia, or excessive menstrual bleeding, affects approximately 22% of healthy women. Due to the failure of drug treatments for the condition, 60% of general practitioner referrals for menorrhagia are treated by hysterectomy within five years. Recently, new minimal access surgical alternatives to hysterectomy, known as conservative treatment, have been developed (Pinion, Parkin, Abramovich, Naki, Alexander, Russell & Kitchener, 1994; Cameron, Mollison, Pinion, Atherton-Naji, Buckingham & Torgerson, 1996).
2.3. Testing for consistency

Following the WTP question, women were presented with a dichotomous choice question which asked them which of the two treatments they preferred. WTP responses were recoded into a nominally scaled variable where \( 1 = WTP_{\text{hysterectomy}} > WTP_{\text{conservative}} \) and \( 2 = WTP_{\text{conservative}} > WTP_{\text{hysterectomy}} \). (Individuals who stated \( WTP_{\text{hysterectomy}} = WTP_{\text{conservative}} \) were not included here.) \( \chi^2 \) tests were carried out to see if the distribution of responses of both the dichotomous choice question and recoded WTP were significantly different across the two alternative treatments (Bland, 1995). The aim here was to establish if the preference structure was significantly different across the two elicitation techniques.

Information on WTP and preferences was then combined to test for consistency of responses. WTP was estimated at the individual level for both hysterectomy and conservative treatment. Individuals should be willing to pay more for their preferred treatment. Responses that did not satisfy this condition were considered ‘inconsistent’. 3 The Fisher Exact Test and McNemar Test were used to test for an association and agreement between the two methods of collecting preference data (Bland, 1995). The aim here was to test whether the implied preference structure was a function of the preference elicitation method.

2.4. Explaining inconsistent answers

Once inconsistent respondents are identified, there is the question of what to do with them. It may be argued that they should be dropped from the analysis since these respondents do not understand the questions or have not taken the exercise seriously. However, Sen (1993) argues that there may be rational reasons for inconsistent responses. In this study an attempt was made to explore the possible reasons for ‘inconsistent’ responses using probit regression techniques. The dependent variable was whether or not the respondent was consistent (\( 0 \) = inconsistent, \( 1 \) = consistent). The independent variables were: responses to the questions asking individuals reasons for their WTP responses; information on ease of completion (indicated by responses

---

3 Responses for which WTP for both treatments were equal were considered ‘consistent’. This was based on the assumption that although some people may prefer one of the treatments, their preference may not be strong enough to pay more for their preferred option.
to the ease of completion question and time to complete the questionnaire) and the professional background of respondents (as a proxy for the information that respondents had about the treatments, and taking a value of “1” for medical staff and “0” otherwise). A general to specific approach was adopted with explanatory variables being excluded by backward elimination if they had a significance level of less than 0.05.

2.5. Implications of including inconsistent respondents

Following this, the influence of including inconsistent responses is considered. WTP is first estimated for the complete sample (including consistent and inconsistent responses), and then just for consistent respondents. It was assumed that WTP was measured on an interval scale, and the t-test was used to see if WTP estimates were significantly different across the two groups. Ninety-five percent confidence intervals were also estimated.

3. Results

There was evidence in the pilot study of inconsistent responses, with individuals willing to pay more for their less preferred option. Analysis of the questions asking individuals the reasons for their answers indicated that respondents were providing cost-based responses. That is, whilst they preferred conservative treatment, they perceived hysterectomy to be more expensive and hence stated a greater WTP for hysterectomy. In an attempt to overcome cost-based WTP values in the main study a distinction was made between cost and value. Individuals were told:

to see why finding out your maximum willingness to pay shows the value you place on something imagine you are going to an auction to buy a car. Having considered all the features of the car (e.g. make, model, colour) you will then decide your maximum willingness to pay (or value) for the car. If you were willing to pay £2000, this is the value to you of the car. If you only pay £1500, then the cost is smaller than the value. Thus, the value you place on the car is not the same as the cost. The same applies to health care interventions. That is, the value may be different from the cost. So, measuring the maximum willingness to pay that people place for some goods is simply a way of measuring the benefits for them derived from the use of these goods.
Of the 146 questionnaires distributed, 75 (51% response rate) were returned. Responses to the WTP and dichotomous choice questions are shown in Table 1. Responses to the WTP question suggest that the majority of respondents prefer hysterectomy, and that the distribution of preferences across the two treatments is significantly different ($\chi^2 = 12.302, P = 0.001$). In contrast, the results from the dichotomous choice question indicate that the majority of respondents prefer conservative treatment. This distribution of preferences across the two treatments is again significantly different, though in the opposite direction ($\chi^2 = 12.552, P = 0.001$). Thus, the WTP and dichotomous choice questions imply different preference structures. These results were confirmed when the WTP and dichotomous choice questions were considered together – 41% of respondents who preferred conservative surgery were willing to pay more for hysterectomy. Both the Fisher Exact Test ($P = 0.018$) and McNemar Test ($P = 0.001$) indicated that the implied preference structure was a function of how preferences are elicited.

A large number of inconsistent respondents gave cost as a reason for their WTP – 60% of inconsistent respondents mentioned cost compared to 17% of consistent respondents. From the regression analysis, cost and ease of completion were significant predictors of inconsistent responses (Table 2). The probability of giving inconsistent answers increased as people mentioned cost as a reason for their WTP answers. People that found the exercise more difficult were more likely to provide consistent answers. This may reflect the fact that those individuals who found the WTP more difficult were more likely to have thought about it. The insignificance of the attribute “occupation” could be due to sample size limitations and to the lack of variation in the attribute since most of the respondents were working at the Maternity Hospital in Aberdeen.

Table 3 compares WTP estimates for the total sample with consistent respondents. For the total sample, those preferring conservative treatment are willing to pay £2031 for conservative treatment and £2290 hysterectomy. Whilst this group are willing to pay more for their least preferred option, the difference is not statistically significant ($t = -0.930; P > 0.15$). Consistent respondents are willing to pay more for the treatment they prefer – those who

---

4 In this study 19 individuals preferred hysterectomy to conservative treatment – five of these were willing to pay equal amounts for hysterectomy and conservative treatment and 14 were willing to pay more for their preferred option. Whilst significantly less of these ‘consistent’ responses stated that ‘cost’ was a reason for their expressed WTP, it is possible that some of these responses were cost-based. The consistency test applied here would not pick this up (since their preferred option was cheaper).
preferred hysterectomy were willing to pay significantly more to have this treatment (£2116 compared to £847; $t = -4.456; P = 0.001$) and those that preferred conservative surgery were willing to pay significantly more for this treatment (£2571 compared to £1711; $t = 2.74; P = 0.001$). Although WTP values are different between the entire sample and consistent respondents, only WTP for hysterectomy for those preferring conservative surgery (as expected, due to all the inconsistencies occurring in this group) is significantly different between the two groups (£2290 compared to 1711; $t = -1.89; P < 0.05$).
4. Discussion and conclusions

This paper found a number of important results when considering the application of WTP experiments to assessing the economic value of a commodity. The results from the WTP experiment were different to those from the dichotomous choice question. This was partly explained by cost-based WTP responses, with 30% of respondents willing to pay less for their preferred alternative. A number of other WTP experiments have found evidence of cost-based WTP responses. Schkade and Payne (1994), in using verbal protocol analysis to investigate how people respond to WTP questions, found that individuals justify their WTP responses by referring to the cost of the commodity being valued. Donaldson et al. (1997) found evidence of cost-based responses in a study looking at different methods of testing for cystic fibrosis. They argued that this problem may arise when open-ended questions are used to elicit WTP values. The study presented here found evidence of cost-based responses when using the PC.

Baron and Maxwell (1996), in a study looking at the value of public goods, provided respondents with different types of costing information. They conclude that future WTP studies should avoid providing respondents with information on costs. The pilot study reported here indicated that even when individuals are not provided with information on costs, they still provide

### Table 3

Mean WTP values for all the sample and the sample of consistent respondents

<table>
<thead>
<tr>
<th>Prefer hysterectomy</th>
<th>Entire group</th>
<th>Consistent respondents</th>
<th>t-Statistic (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean WTP</td>
<td>2100</td>
<td>2116</td>
<td>0.77</td>
</tr>
<tr>
<td>Hysterectomy (£) (95% CI)</td>
<td>(1465.6; 2734.3)</td>
<td>(1701.8; 2956.9)</td>
<td>(P &gt; 0.15)</td>
</tr>
<tr>
<td>Mean WTP</td>
<td>1031</td>
<td>847</td>
<td>-0.48</td>
</tr>
<tr>
<td>Conservative (£) (95% CI)</td>
<td>(487; 1574.9)</td>
<td>(388.3; 1435)</td>
<td>(P &gt; 0.25)</td>
</tr>
<tr>
<td>Strength of preference</td>
<td>2.03</td>
<td>2.55</td>
<td></td>
</tr>
<tr>
<td>t-Statistic (P-value)</td>
<td>t = -3.217, P = 0.005</td>
<td>t = -4.456, P = 0.001</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prefer conservative</th>
<th>Entire group</th>
<th>Consistent respondents</th>
<th>t-Statistic (P-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean WTP</td>
<td>2290</td>
<td>1711</td>
<td>-1.89**</td>
</tr>
<tr>
<td>Hysterectomy (£) (95% CI)</td>
<td>(1625.6; 2954.3)</td>
<td>(1129.9; 2334.3)</td>
<td>(P &lt; 0.05)</td>
</tr>
<tr>
<td>Mean WTP</td>
<td>2031</td>
<td>2571</td>
<td>0.99</td>
</tr>
<tr>
<td>Conservative (£) (95% CI)</td>
<td>(1492.2; 2569.7)</td>
<td>(1629.1; 3199.4)</td>
<td>(P &gt; 0.15)</td>
</tr>
<tr>
<td>Strength of preference</td>
<td>0.88</td>
<td>1.39</td>
<td></td>
</tr>
<tr>
<td>t-Statistic (P-value)</td>
<td>t = -0.930, P &gt; 0.15</td>
<td>t = 2.74, P = 0.001</td>
<td></td>
</tr>
</tbody>
</table>

* The values forming the ratio are significantly different at the 95% level.
** Significantly different WTP values between the entire sample and the sample of consistent respondents at the 95% level.
cost-based responses. Following this, the main study reported here directly attempted to avoid cost-based responses by including a statement in the questionnaire explaining that cost is not the same as value. Despite this, the experiment still found evidence of cost-based responses. Further, such an approach may have the added disadvantage that highlighting the cost-value distinction may alert individuals to the concept of cost.

Questions are raised concerning the effect of cost-based WTP on the application of WTP experiments within a cost–benefit analysis (CBA) framework. If we assume that inconsistent respondents have not understood the task and hence have not answered in terms of their maximum WTP, their responses do not reflect their maximum value but rather the cost of treatment. If cost is being used as a marker for expressing WTP then WTP for low cost options may be underestimated and WTP for high cost options may be overestimated. Inefficient policy conclusions may therefore result.

Given that cost-based responses will lead to inaccurate estimates of welfare changes, how may such responses be avoided? To date it has been shown that both the open-ended and PC WTP techniques are vulnerable to such responses. Would the referendum style WTP approach also be subject to such biases? This method was proposed by the US National Oceanographic and Atmospheric Administration (NOAA) Panel on contingent valuation as the method of choice in WTP studies of the environment (Arrow, Solow, Portney, Leamer, Radner & Schuman, 1993). This recommendation was based on the premise that this technique would not be subject to biases, though this argument has since been challenged (Diamond & Hausman, 1994). Cost-based responses may still result from the referendum approach. For example, if respondents think the commodity being valued costs less than the bid amount they are presented with, they may be more likely to respond negatively to the bid amount presented to them. Similarly, if they think the commodity being valued costs more than the bid amount they are presented with, they may be more likely to respond positively to the bid amount presented to them. This behaviour would be consistent with the “fair price” explanation for WTP responses, i.e. people like to pay what they believe it

---

5 Using this approach, individuals are asked whether they would pay a specified amount for a given commodity, with possible responses usually being ‘yes’ or ‘no’. The bid amount is varied across respondents and the only information obtained from each individual respondent is whether her maximum WTP is above or below the bid offered (Hanemann, 1984).

6 The NOAA panel was a group of “experts” who offered guidelines on the use of WTP for environmental resource damage assessments.
costs. They do not want to be exploited by paying more than the commodity costs, nor to exploit by paying less than the commodity costs. There is evidence of this ‘fair-price’ behaviour in the provision of both private and public goods (Kahneman, Knetsch & Thaler, 1986; Winer, 1986; Green, Kahneman & Kunreuther, 1994).

One way to attempt to overcome cost-based responses is to focus on the benefits from the commodities being valued (Baron & Maxwell, 1996). The marginal WTP approach may be used to do this (Donaldson et al., 1997; Donaldson, Hundley & Mapp, 1998). Here individuals are first asked what option or treatment they prefer. They are then asked their maximum WTP to have their preferred option over their less preferred. A second alternative is the use of conjoint analysis (CA) discrete choice models to estimate WTP. Using CA respondents are presented with numerous discrete choices, which differ in terms of the attributes of the service, and, for each choice, they are asked to choose their most preferred. By estimating a difference model, and including cost as an attribute, marginal WTP values for the change from one treatment to another can be obtained (Ryan & Hughes, 1997). Future work should explore whether cost-based responses emerge when these instruments are employed.

Even if the potential problem of cost-based responding in WTP experiments is successfully overcome, research has indicated a number of other problems with using this technique to elicit the economic value of a good. These include the insensitivity of WTP estimates to factors that might be expected to influence such responses (Kahneman, 1986; Kahneman & Knetsch, 1992; Diamond & Hausman, 1994; Irwin & Spira, 1997); the sensitivity of WTP estimates to factors which are theoretically irrelevant to the commodity being valued (Bishop, Heberlein & Kealy, 1983; Sellar, Chavas & Stoll, 1985; Cummings et al., 1986; Peterson, Driver & Gregory, 1988; Johnson, Bregenzer & Shelby, 1990; Samples & Holier, 1990; McFadden & Leonard, 1992; Boyle, Johnson, McCollum, Desvousges, Dunford & Hudson, 1993; Irwin, Slovic, Lichenstein & Mcelland, 1993; Kealy & Turner, 1993; McFadden, 1994; Ready, Buzby & Hu, 1996; Kramer & Mercer, 1997); the presence of strategic behaviour (Bohm, 1972; Cummings et al., 1986; Mitchell & Carson, 1989; Posavac, 1998); and the extent to which values exist for unfamiliar goods (Fischoff, 1991; Payne & Bettman, 1993).

In conclusion, this paper developed a method to test for the consistency of responses in WTP experiments – if commodity A is preferred to B, then individuals should be willing to pay more for A than B. Thirty percent of respondents fail to provide consistent responses. One reason for such
inconsistencies was cost-based responses. These findings are consistent with other studies in the WTP literature. The results are important when considering the application of WTP experiments both in health care, and elsewhere. Whilst economists argue the case for the use of WTP on the grounds that the instrument provides a cardinal measure of preferences, the evidence presented here suggests that the technique may not even provide an ordinal measure of preferences. Possible solutions to avoid cost-based WTP responses include the marginal WTP approach and CA. Future work should explore the consistency of these approaches, as well as the validity of WTP experiments more generally.

Acknowledgements

The authors would like to thank Emma McIntosh and two anonymous referees for comments on this paper. Financial support from the Medical Research Council (MRC), the Chief Scientist Office of the Scottish Office Department of Health (SODOH), and the government of Navarre are acknowledged. The views expressed in the paper are those of the authors.

References


