Contingent Valuation:
Indiscretion in the Adoption of
Discrete Choice Question Formats?

Richard D Smith
Lecturer, Health Economics Unit, CHPE

October, 1997
ISSN 1325-0663
ISBN 1 875677 83 6
CENTRE PROFILE

The Centre for Health Program Evaluation (CHPE) is a research and teaching organisation established in 1990 to:

- undertake academic and applied research into health programs, health systems and current policy issues;
- develop appropriate evaluation methodologies; and
- promote the teaching of health economics and health program evaluation, in order to increase the supply of trained specialists and to improve the level of understanding in the health community.

The Centre comprises two independent research units, the Health Economics Unit (HEU) which is part of the Faculty of Business and Economics at Monash University, and the Program Evaluation Unit (PEU) which is part of the Department of Public Health and Community Medicine at The University of Melbourne. The two units undertake their own individual work programs as well as collaborative research and teaching activities.

PUBLICATIONS

The views expressed in Centre publications are those of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of publications are encouraged to contact the author(s) with comments, criticisms and suggestions.

A list of the Centre's papers is provided inside the back cover. Further information and copies of the papers may be obtained by contacting:

The Co-ordinator
Centre for Health Program Evaluation
PO Box 477
West Heidelberg Vic 3081, Australia

Telephone +61 3 9496 4433/4434 Facsimile +61 3 9496 4424
E-mail CHPE@BusEco.monash.edu.au
ACKNOWLEDGMENTS

The Health Economics Unit of the CHPE receives core funding from the National Health and Medical Research Council and Monash University.

The Program Evaluation Unit of the CHPE is supported by The University of Melbourne.

Both units obtain supplementary funding through national competitive grants and contract research.

The research described in this paper is made possible through the support of these bodies.

AUTHOR(S) ACKNOWLEDGMENTS

The author would like to express his thanks to Associate Professor Jan Abel Olsen, of the University of Tromso, for his comments on earlier drafts of this paper. Needless to say, the usual disclaimer applies.
Contingent valuation (CV) refers to a hypothetical survey method of valuing the benefits of an intervention in monetary terms by estimation of the individuals maximum willingness-to-pay (WTP). Currently, an issue in the application of CV methods is the technique used to elicit this monetary valuation. Historically, the favoured technique has been the ‘open-ended’ questionnaire, where the respondent is asked directly for their maximum WTP for the commodity being valued. However, in recent years there has been a move away from the use of this open-ended technique, towards the use of discrete choice questionnaires (also referred to variously as closed-ended, binary or dichotomous choice questionnaires; as well as referendum surveys if the median, rather than mean, WTP is the desired measure of value). In contrast to open-ended questionnaires, discrete choice questions offer the respondent a single value (bid), which they either accept or reject. By varying this single bid across various sub-samples a demand curve for the commodity is estimated, and from this the maximum WTP calculated.

The basis for the use of discrete surveys in preference to open-ended seems to rest on some combination of: (i) a belief that this technique offers a more ‘realistic’ market, and will therefore lead to more valid responses (a truer estimate of actual WTP by respondents); and (ii) the supposed tendency of discrete surveys to yield higher response rates, through reduced mental demands (especially for mailed surveys). This paper presents a review of the use of discrete versus open-ended survey techniques, and addresses these issues as well as others of importance.

It is concluded that, although there are issues yet to be resolved concerning the degree of bias within the open-ended technique, there appear to be substantial additional issues with the use of discrete survey techniques. While the discrete choice questionnaire seems to have been favoured, major problems still remain in its implementation. These problems arise from the fact that, simply, discrete choice survey design and analysis is very complex. Survey design elements include total sample size, bid range, specific bid levels, allocation of the total sample among the bid levels, and form of statistical model used to analyse such data. None of these issues has by any means been clearly resolved.

This author would therefore suggest that the choice of discrete versus open-ended techniques is by no means settled, and is potentially a red-herring in the search for ‘valid’ WTP values derived from CV surveys. There appears no reason why the open-ended survey should be summarily dismissed, and research may better be targeted to refinement of the open-ended approach to reduce bias. In the meantime, it is recommended that caution be exercised in the seemingly indiscreet adoption of the discrete choice question approach.
# TABLE OF CONTENTS

1   Introduction    1

   1.1  NOAA Panel Recommendations  2

   1.2  Aim of the Paper  2

2  Open-Ended Versus Discrete Questionnaire Design  3

   2.1  Developments of Discrete Questions: Double and Triple Bounding  5

3  Issues in Assessing Discrete Versus Open Questionnaires  7

   3.1  Response Rate  7

   3.2  Market Realism and the Potential for Strategic Behaviour  8

   3.3  Starting Point, Range and ‘Yea-saying’ Bias  11

   3.4  Bid Vector to be used in Discrete Surveys  14

   3.5  Statistical analyses Required to Calculate Mean WTP  15

      3.5.1  Model Specification and Functional Form  16

      3.5.2  Confidence Intervals  18

   3.6  Use of median Versus mean Estimate of WTP  19

4  Discussion  22

References  24
1 Introduction

Contingent valuation (CV) refers to a method of valuing the benefits of an intervention in monetary terms which is not based on actual observation of a market value, or its proxy (such methods commonly referred to as “revealed preference” studies) (1). Rather, it is the valuation of preferences, expressed as maximum willingness-to-pay (WTP), or minimum willingness-to-accept compensation (WTA), as elicited from hypothetical surveys. Such surveys may be administered either by interviewer (in person or by telephone) or mail. The CV method has been principally developed in environmental and transport economics (2-5), but over the last decade has made considerable headway in the economic evaluation of health care interventions (6-19).

Currently, an issue in the application of CV methods is the technique used to elicit this monetary valuation (generally WTP, rather than WTA). Historically, the favoured technique has been the ‘open-ended’ questionnaire, where the respondent is asked directly for their maximum WTP for the commodity being valued. Variations in this method have been developed to ensure higher response rates and more reliable and ‘valid’ responses, such as bidding games and payment cards (see section 2). However, in recent years there has been a move away from the use of this open-ended technique, towards the use of discrete choice questionnaires (also referred to variously as closed-ended, binary or dichotomous choice questionnaires; as well as referendum surveys if the median, rather than mean, WTP is the desired measure of value) (1). In contrast to open-ended questionnaires, discrete choice questions offer the respondent a single value (bid), which they either accept or reject. By varying this single bid across various sub-samples a demand curve for the commodity is estimated, and from this the maximum WTP calculated (1).

As with CV in general, this discrete choice method of eliciting WTP has been developed in the environmental economics field over the last 10 years, with gaining popularity and momentum (20-24). Perhaps the greatest impetus to its widescale adoption has been the recent influential recommendations of the National Oceanographic and Atmospheric Administration (NOAA) in the USA concerning the conduct of WTP studies for the valuation of environmental benefits (20).

1.1 NOAA Panel Recommendations
Following the Exxon Valdez oil spill in Alaska, and subsequent debate over the use of CV in estimating damages to be awarded against Exxon, the NOAA commissioned a panel of leading figures in the use of CV in valuing environmental benefits to assess the validity of CV methods (20). The implications of this report for evaluations of CV in healthcare have been raised (6, 25). However, for the purposes of this paper, there were two important recommendations made by the NOAA panel. First, that discrete choice questions should be used in preference to open-ended questions. The NOAA panel concluded that responses to open-ended questions were ‘biased, erratic and unreliable’, because of being both difficult to answer and unrealistic. The panel felt that the discrete format was more likely to lead to ‘valid’ WTP responses. Additionally, the panel also recommended that only one such question be presented to the respondent (this becomes relevant in section 2.1, in the discussion of single bounded discrete questions versus double and triple bounded). The second recommendation was that face-to-face interviews should be used, with interview by telephone acceptable in some cases. That is, mail out questionnaires were not recommended. Again, this has implications for adoption of discrete questionnaires which are outlined in section 3.1.

Although the NOAA panel’s recommendations have received little critical attention, Donaldson et al (6), in a recent paper, put forward an argument that, although the results of their study were felt to not reflect maximal WTP (due to a focus on the ‘cost’ of the intervention), this did not “necessarily lead to the conclusion of the NOAA Panel that questions be asked in a referendum format” (p154/5). Donaldson et al (6) then proceed to suggest other means to cut down on ‘invalid’ responses which do not use such discrete questionnaire formats. However, such critical examination of the NOAA panels recommendations have been scarce and the discrete question format is beginning to dominate the CV of health care interventions as it has with environmental commodities.

1.2 Aim of the Paper

This paper therefore seeks to critically reassess the apparent benefits of the discrete choice format over the open-ended question format as a means to elicit WTP in CV studies of healthcare interventions. A brief description of the two methods, with the reasons behind a move from open-ended to discrete questionnaires, is provided in section 2. Section 3 then critically surveys the major issues concerned in assessing the relative superiority of the discrete choice, compared to open-ended, format, including consideration of response rates, strategic behaviour, starting point bias, ‘bid vector’ required in discrete surveys, and issues in the statistical analysis required to estimate mean WTP. The paper concludes with a discussion and recommendations in section 4.
2 Open-Ended Versus Discrete Questionnaire Design

The essential difference between open-ended and discrete questionnaire formats is, in theory, simple. In open-ended valuation questions the researcher may use interviews (face-to-face or phone) or mailed questionnaires to elicit directly the respondents maximum WTP. In contrast, in discrete valuation questions the respondent either accepts or rejects a single value (bid) which is presented to them (1).

For open-ended questions some form of visual aid is usually used to assist the respondent in making their valuation. These aids are felt necessary due to demonstrated poor response without such assistance (1, 3). One aid of particular popularity is the “bidding-game”, which, as its name suggests, resembles an auction process. In this case a bid is made to the respondent, who then accepts or rejects it. The next bid offered is therefore higher or lower, as appropriate, and this process continues until the respondents maximum WTP is reached. Another popular form of aid is the “payment card”, where the respondent is presented with a range of values from which to choose (1, 6).

However, there are two problems which have been encountered in using open-ended questionnaires. First, problems associated with using visual aids, particularly starting point and range bias\(^1\) (1). Second, low response rate for mail based surveys (although response to face-to-face interviews is predictably higher) (1). It is, in the main, these two problems that led to the development of the discrete choice approach.

Discrete questions differ from open-ended questions by offering the respondent only one bid, or value, which they either accept or reject. Because each respondent only provides this limited information (that their true maximum WTP is either equal, greater or less than the specified bid), the critical feature of such a survey is the requirement for different sub-samples to be offered different bids. By varying the bid across different sub-samples in this way it is possible to calculate the proportion of respondents who are willing to pay as a function of the bid. The resultant ‘curve’ is then interpreted as the aggregate demand curve for the commodity valued, if the proportion responding positively to the bid is multiplied by the number of respondents. In simple terms, if we assume that 100% of the sample will pay a zero bid, as the bid increases from zero the proportion willing to pay that amount will fall, up to that bid where no-one (0% of the sample) is willing to pay it. Analogous to the theory of demand, the area under the curve is then taken as the societal WTP. In some cases a ‘referendum’ approach is favoured, whereby the bid at which 50% of the sample are WTP and 50% not WTP, the median, is used as the point at which the majority of population will benefit (the use of this median approach, versus the mean WTP, is discussed in section 3.6).

\(^1\) For further discussion of these, and other, biases see Mitchell and Carson (26).
Two advantages are suggested for this discrete choice approach. First, it is argued that it closely resembles a market situation, where price is given, which contributes to greater respondent understanding, and therefore yields more 'valid' WTP values, because of its realistic representation of the consumers common choice in the market. In addition, the use of one such question also ensures a high response rate, as well as providing less room for strategic behaviour (26, 27). Second, it avoids starting point, and/or range, bias in the use of aids for open-ended questions, since individuals are only presented with one bid which they either accept or reject.

However, the discrete choice format also has potential drawbacks. First, the information derived from each respondent is minimal compared with that from an open-ended questionnaire. That is, for open-ended questionnaires the respondent states exactly their maximum WTP, which is the desired figure. However, for discrete questionnaires the respondent only portrays whether or not they are willing to pay a given value (or price). Thus their maximum WTP is not directly provided, and we can only infer that it lies somewhere (i) equal to or above the specified value, or (ii) equal to or less than the specified value. In order to use these values, statistical analyses are required to estimate the ‘demand’ curve, from which the sample mean (or median) can be derived (1). Since little information is gained from each respondent, discrete surveys therefore require a large sample size: far in excess of that required for an open ended WTP survey. For example, it has been suggested that this sample may approach 1,000, or even greater (28, 29), which would involve significant research monies.

Second, in order to estimate this mean (or median) from discrete responses, some form of regression, or non-parametric statistical analysis, must be used. The most popular technique to analyse such data is logistic regression (1). Here the probability of acceptance of a bid is calculated as a function of the bid. Since maximum WTP is uncertain using this technique, often the analysis is truncated at the maximum bid. Similarly often the analysis is truncated at zero rather than allowing for negative bids, although it has been suggested that these may be of some importance (30). Not only are fairly complex statistical procedures required, but also, in the case of parametric approaches, distributional assumptions are required which can greatly affect the resultant WTP estimate (31). Neither of these is necessary with analysis of open-ended WTP.

However, in an attempt to circumvent some of these problems, such as providing greater information by narrowing the range where the respondents ‘true’ WTP lies, discrete choice questionnaires have been developed which offer the respondent more than one choice, based on their response the preceding choice. This is an important development, and is discussed in more detail below.

---

2 Such surveys are named according to how many of these discrete choices are made, typically being double or triple ‘bounded’ discrete choice valuation questionnaires (where the initial discrete choice used in isolation is termed single bounded).
2.1 Developments of Discrete Questions: Double and Triple Bounding

As mentioned, although discrete choice questionnaires have gained popularity over the last few years, due primarily to their purported advantages in avoiding bias associated with open ended CV questions, they are highly statistically inefficient in that a very large number of observations is required to identify a distribution of values with any degree of accuracy (32). An alternative strategy, intended to reduce this inefficiency (and therefore reduce the large sample sizes required), was first proposed by Hanemann (33) and implemented by Carson et al (34). In this approach a sequence of questions is used to narrow the range within which the respondents ‘true’ WTP is postulated to lie. The respondent is presented with an initial bid, bid 1, which they either accept or reject. Following this, they are offered a second discrete response dependent upon their response to the first. In general, if a respondent indicates a willingness to pay the first offered amount the second bid choice will be offered at an amount approximately double the first, bid 2; if they are unwilling to pay this first bid then this second bid is set at approximately half the initial level, bid 3\(^3\) (21). This sequence will thus yield the individuals value in one of four possible regions; zero to bid 3, bid 3 to bid 1, bid 1 to bid 2 or above bid 2 (where we may assume that bids will be bounded by the individuals income level). The result is a significant increase the amount of information provided by each individual on the bounds within which their ‘true’ WTP is postulated to fall, which has been shown to be asymptotically more efficient than single bounded surveys (21).

However, the immediate issue with this approach is moving away from one single bid creates a situation analogous to the bidding game used in open-ended surveys. The only difference is in truncating the bidding game at a point earlier than when the respondent gives their maximum WTP (ie double/triple bounded discrete surveys may be considered to be bidding games with only two/three bids). In this way, one may assume that the double/triple bounded discrete survey would be subject to the same potential for starting point and strategic bias as the open-ended bidding game, but at the same time providing less information. This was recognised by the NOAA panel, for example, who, although recommending the use of discrete choice surveys, note parenthetically that “if a double bounded dichotomous choice or some other question form is used to obtain more information per respondent, experiments should be developed in order to investigate biases that may be introduced” (35, p52). Needless to say, little research has been conducted (32). Such double, or triple, bounding would therefore seem then to potentially offer the worst of both worlds: the biases argued to be inherent in the open-ended approach combined with the limited data provided by the discrete choice approach. To favour the double/triple bounded approach is to implicitly state that starting point bias is less of a problem with only two/three bids than with more than two/three which does not seem to be supported by available evidence (32, 36, 37). This issue is again considered in section 3.3.

Overall, the discrete choice question format is proposed as a means to increase response rates, especially for mailed questionnaires, remove potential for starting point bias and reduce the incentive and possibility for strategic bias. However, as suggested, the issues are not so clear-

\(^3\) For triple bounded surveys a third discrete choice is similarly given based on the answer to the second.
cut, and there are additional problems peculiar to discrete choice questions which may negate the potential benefits, such as requirement for large sample size, design of appropriate bid vector and degree of statistical analysis required to estimate mean WTP. It is to a discussion of these issues in more detail that the paper now turns. Before that, it is worth noting that other possible biases in CV surveys, such as scenario misspecification, will be equally likely to be present in both open-ended and discrete question formats. These issues are therefore not considered in this paper, and the interested reader is referred to Mitchell and Carson (26).
3 Issues in Assessing Discrete Versus Open Questionnaires

3.1 Response Rate

Non-response causes problems in any survey if the sample not responding is likely to be systematically different from those responding. That is, in the case of contingent valuation if the non-responders would have a significantly different WTP compared to those who did respond. Ensuring a high response rate is therefore important for any survey, for this reason as well as to ensure survey expenditure is kept to a minimum. Considerable efforts are therefore expended in trying to ensure a survey will yield as high a response rate as possible. For example, this was a major reason behind the development of the bidding and payment card approaches to assisting respondents in answering open-ended contingent valuation surveys.

The response rate, however, is a complex function of many factors. These include the overall level of difficulty of the questions posed, the amount of information provided which may lead to cognitive overload in response, the physical length of the survey leading to respondent fatigue, the use of personal face-to-face, or telephone, interviews versus postal questionnaires and the use of financial or other incentives to complete, such as entry in prize draw or promise of lottery tickets (38).

In terms of response rate, the debate between open-ended and discrete choice surveys is simple. By being shorter and more ‘realistic’ the discrete choice survey is argued to be simpler and more likely to yield a higher response rate, particularly in postal surveys (which are considered desirable as they are less costly than personal interview surveys). There are two issues with respect to this proposition.

First, is it true that discrete choice surveys yield higher response rates? Given the complex mix of factors determining response rate, as indicated, we would expect that, ceterus paribus, making the questionnaire shorter and possibly simpler would increase response rate, and in some cases this has been shown (8), although in others response rates have been equal (39). However, when we relax the ceterus paribus assumption, the case is not so clear. For example, it was reported that in a direct comparison of discrete choice versus open-ended payment card mailout surveys, that response was higher by 2% for the payment card survey, and also that protest bids were greater for the discrete choice survey than the payment card approach (40). One may expect response rate to depend heavily on whether it is mailed or interviewer, and incentives for completion. For example, one recent study using discrete choice questionnaires had a high response rate, but this was at the expense of a $4 payment and 3 follow-up mailouts (36). Furthermore, one recent study of open-ended versus discrete mailout questionnaire found a response rate for open-ended of 85%, whilst response rate for discrete was 73% (41).

It is also important to note that double and triple bounded surveys have typically been used in phone or personal surveys, where the flow of questions is controlled by the interviewer. The use
of these in mailout surveys raises similar issues to that of postal bidding surveys of the ability and willingness to respondents to follow the inevitably more complex survey instructions. For example, a 5% to 7% reduction in response from the addition of second discrete question (ie double bound) has been found (37).

A second issue is whether the use of discrete choice surveys, if we assume that they do indeed lead to higher response rates, will be cheaper to conduct than open-ended surveys. This premise is based on discrete choice surveys yielding high response rates in postal settings, compared to poor response to open-ended surveys. There are two issues here. First, let us assume that this is correct, and further that discrete choice surveys yield a response rate of 100%, but open-ended a response rate of 50%. In this case, if the discrete survey requires a sample size of, say 1,000 (as indicated earlier), and the open ended a sample of 400, then surely the open ended will be cheaper, as it only requires a survey of 800 people to get its required sample of 400. This of course assumes that the systematic problem of non-response is not an issue. Let us therefore, second, assume that this postal response rate is unacceptable for open-ended questionnaires and so telephone interviews are required, which yield a response rate of 80%. In this case, we are looking at the comparative cost of 1,000 mailed questionnaires, versus 500 telephone interviews which may well be approximately identical, or again the open-ended survey could be slightly cheaper.

A final point worth noting is the inconsistency created by arguing for the superiority of discrete choice questionnaires based on the response rate to postal surveys. If we are using the discrete survey because of NOAA recommendations of its greater validity we should not, to be consistent, be using them in mail out surveys at all as this violates the NOAA recommendation mentioned earlier of using face to face, or at the very least telephone, surveys. The NOAA may explicitly reject the use of open-ended surveys, but they also explicitly reject the use of mailout surveys.

However, perhaps response rate may be considered a minor, even if important, issue in assessing whether discrete or open-ended questionnaire designs are to be favoured. In this case let us move on to discuss an arguably more substantive issue: that the discrete survey design will yield more ‘valid’ estimates of WTP.

3.2 Market Realism and the Potential for Strategic Behaviour

It has been reported that, where open-ended and discrete questionnaires have been compared, substantial disparities are found between the results (39, 41-46). For example, mean and median values of around two to three times higher from discrete versus open-ended surveys (42, 43, 46), although some suggest no significant difference (44). However, a recent meta-analysis of a large number of CV studies found substantial evidence supporting the hypothesis that discrete choice surveys tend to give higher mean values than open-ended surveys. Various reasons are postulated for this, in particular: (i) incentives for strategic behaviour differ; and (ii) ability and willingness to cooperate with the method. For example, Mitchell and Carson conclude that the discrete method “is considered easier for respondents to answer than other elicitation formats that use payment card or open-ended questions” ((26), p379), whilst Hoehn and Randall claim that
discrete questions are less vulnerable to strategic bias (27). Such comparative studies are important because, if the WTP values are found to be sensitive to the question format used, then the validity of either one, or both, is questioned. It would then be imperative to establish under what circumstances one of the question formats has a distinct advantage. It is important, though, to recognise that economic theory can provide no reason for such disparity. The important question, then, if such disparities do occur is why discrete should be preferred - why are the responses to these questionnaire seen as more valid than open ended CV? This is the critical question for research at present (39). One answer may lie in the supposed ‘realism’ of the discrete versus open-ended technique, and the limited scope for strategic behaviour (42).

The trade off in any contingent valuation survey is between making the market of sufficient realism to elicit a true valuation, but not making it so realistic as to provide considerable incentive for strategic bias. It has been argued that discrete choice surveys will meet this trade off more effectively than open-ended surveys because: (i) the choice facing the respondents is close to the market situation they commonly deal with, and will therefore be more realistic; (ii) and providing them only one value to accept or reject almost totally removes the potential for strategic behaviour, even if the respondent wished to behave in such a way. This section considers each of these in turn.

First, the market mechanism is more realistic. It is argued that in open-ended surveys the respondent is faced with a highly unusual, and therefore unfamiliar, valuation mechanism. In such cases they may either be unable to provide a response, or search for some indication of what the value ‘should’ be. For example, they may try to estimate the approximate cost and base their valuation on this (6). That is, in the absence of a market price the respondent has to ‘search’ for the valuation, which is time consuming and mentally tiring. To expedite this search they may either simply not attempt to search or may search for a near proxy, such as the ‘cost’. In contrast, with discrete choice surveys the respondent’s answers may be more ‘valid’ because of presenting the respondent with a common and familiar ‘market’ choice of accepting the bid or not. ie purchasing the good at the stated price or not.

However, this is not the only means by which the respondent may be presented with a realistic market scenario. For example, it could be argued that “individuals ‘shop around’ observing different prices for a good and, therefore, that, as well as closed-ended questions, payment scales are a relevant alternative to the open-ended approach” (6, p155). That is, if the use of discrete questions is argued to be to provide the respondent with a guide to price, and they therefore simply have to refer to whether the good in question provides value at least equivalent to that price, then there are other mechanism to provide this guidance.

The essential point is that respondents will be searching for their valuation in either survey design, and will possibly require assistance to discover this valuation. If they wish to use an implicit estimation of what the likely ‘cost’ of the commodity in question is, then this idea of cost could be behind open and discrete answers, in providing maximal WTP equivalent to an estimate

---

4 Although the reliability of both, at least, may be equivalent and high (around 0.6 test-retest correlation) (47).
of the cost in open-ended surveys, or using this estimate to decide whether or not to respond yes to a bid in discrete surveys. For example, if the respondent estimates that the cost would be $50, then this may translate as a either a valuation of $50 in an open-ended question, or a response of yes to bid $50, but no to bid of $51 in a discrete survey. The fundamental concern here is what is it that influences WTP, not what will influence responses to open versus closed questions. In addition, there is still equivocal evidence of the ‘validity’ of open-ended versus discrete questionnaires. There are studies which have found that discrete questionnaires provide less non-response or zero (often inferred as protest) bids (7, 48). However, there are also studies which show that open ended questionnaires are ‘valid’ (49).

Second, it is argued that the opportunity for strategic behaviour in discrete choice surveys is minimal (27). In essence, strategic bias refers to the respondent feeling that their interests will best be served by giving a higher or lower value than their true WTP. For example, if it is thought by the respondent that the provision of the commodity depends on their valuation, but that payment is independent of the amount they express, then the incentive is to exaggerate their WTP, and vice-versa. Strategic bias, and the potential of the survey to minimise it, have been considered at length in the use of contingent valuation (26). Strategic behaviour therefore depends critically on the interaction of two factors: (i) the likelihood with which respondents feel that the WTP value given will be charged to them; and (ii) the likelihood that the WTP value given will determine the provision of the good. These motives may apply equally to open-ended or discrete surveys, which is not, to this author’s knowledge, in dispute. The issue is one of if the respondent wished to behave in a strategic manner they could have a large impact on group WTP using the open-ended survey, but, in contrast, with discrete surveys, it is argued that since only the acceptance or rejection of one specified bid is required, it is difficult for any one individual to have a significant effect on the overall mean WTP (compared to an open survey). However, although it is difficult for any one individual to have an effect on the overall mean WTP, compared to an open-ended survey, if many or all respondents answered discrete questions in a strategic manner then results of discrete and open questions would be affected to same extent (1).

Empirically, strategic bias has never been clearly demonstrated to be present to any significant extent (32, 50-55), and that if it is present it’s effect is likely to be minor (56, 57) (although it is possible that strategic behaviour may vary according to what proportion of a respondents income the commodity may require, suggesting the possibility of ‘thresholds’ in strategic behaviour (58)). For example, it has been noted that “open ended does not necessarily cause problems of protest answers and difficulties in answering questions” (30, p 100). Other authors though, such as Cameron (59), argue that the dichotomous nature of the discrete question format “circumvents much of the potential for strategic bias” (59, p355).

However, it has been shown that strategic behaviour does not necessarily vary across discrete or open-ended methods (60), although “greater consideration should be given to the identification and treatment of respondent uncertainty in CVM surveys and, in particular, with closed-ended formats” (60, p306). That is, in a discrete survey respondents are more likely to abstain than accept a value which overstates their bid, thus biasing results downward compared with open-ended questionnaires.
A final point worth noting is that there may well be a distinction to be made between the ‘validity’ of open-ended surveys in environmental commodities, and health care interventions. A recent study, which tested for different results in public good versus private good valuation, found no difference between open-ended and discrete values for the private good, only the public good (42). This, the authors suggest, may be because the incentive for strategic behaviour may not exist for private goods. Private goods tend to be concrete and well-defined and so it is less likely that the respondent will face difficulty in valuation of these, whatever the method. The authors do, though, qualify this by outlining that different payment mechanisms may produce different results, such as payment via taxation versus out-of-pocket and so we cannot conclude that the discrete survey is the more valid. Rather it means that there needs to be more theoretical work on models underlying these techniques to explain differences and which is therefore expected to be more valid. Of particular importance here is the implication for the valuation of health care interventions. One could argue that we may expect values to differ less, compared to valuing environmental commodities, by use of discrete versus open-ended survey design since many of these interventions are more ‘private’ than ‘public’ in nature (ie closer to Mars bar than a rain forest).

In general we may conclude that the degree of realism with which the market and payment is perceived by the respondent is likely to be similar across open-ended and discrete surveys. In addition, incentives for strategic behaviour are also identical, with the only difference of any significance being that in single bounded discrete surveys the opportunity for any one individual to influence the mean WTP is less than with open-ended surveys. However, this advantage is negated as soon as we move toward double and triple bounded surveys, and the potential for all respondents, or sub-groups, to act strategically in unison.

3.3 Starting Point, Range and ‘Yea-saying’ Bias

As mentioned, traditional open-ended surveys have tended to commonly use one of two techniques for improving the response to questions; bidding or payment cards. Both of these are forms of prompts to the respondent providing them with: (i) a starting bid which is then iteratively accepted or rejected until a final WTP amount established; (ii) a range of values from which to ‘choose’ their WTP. It has been argued that these measures therefore create a bias in the form of first response or range bias (61). It is the former which is of interest here, and the degree to which the potential for it may differ between open-ended and discrete questionnaires. Within the general classification of first response bias there are two forms which are of most interest here: starting point bias and yea saying, and these will be considered in turn.

First, starting point bias. The bidding approach has long been vilified for being subject to possible starting point bias (26, 62). In particular the concern that an initial prompt or bid in the bidding approach will anchor the respondent. The idea is that a respondent has largely uncertain views on the value of the product presented, and so forms WTP valuations based on prior experience, commodities which are ‘close’ and underlying beliefs. However, respondents, in wishing to have as much information as possible on which to base their value, may view the bid
or prompt offered as representing what ‘society’ or ‘experts’ view as the value of the good, or what it may cost (6, 61). The problem of starting point bias is that it will bias the mean value toward the starting bid, narrow the distribution around the mean (portraying greater consensus than truly exists), and causes a loss in efficiency. It is important to note, though, that there are instances where starting point bias has been tested for and not found (39).

In contrast, it is argued that with discrete questions, since there is only one bid which is accepted or rejected, and these bids are varied across subsamples, there is no possibility of such starting point bias. With single bound questionnaires this may be the case, but as we have seen there is a growing move to the use of double, or triple, bounded questionnaire (see section 2.1).

Whilst efficiency may be improved, several authors (2, 21) have found that the WTP estimates yielded by double bound are significantly different than those from single bound. There are various reasons this may occur (21, 22, 28), but there is some evidence that starting point bias may be the cause (36, 37). This seems intuitive, since double, or triple, bounded surveys may be seen as merely a limited (and possibly more structured) version of the more traditional bidding approach to CV (21). The respondent, when answering the second discrete question, no longer compares their ‘true’ WTP to the follow up bid, but combines their ‘true’ WTP with the initial bid to form a revised WTP which is then compared to the follow up bid. Such bias has been found to be a potential factor in analysis (37).

The implicit assumption of the double (triple) bound discrete approach is that the respondents answers to both (all) of the questions are driven by one underlying WTP value (24, 32). If this is true then the second (third) question increases information about the true WTP contained in the answer because it creates a tighter interval around this true WTP. However, if this is not true, and the two (three) values are definitely not identical (although highly correlated), then assuming they are can severely distort the estimated value distribution (32). The researcher is required, therefore, to explicitly recognise the endogeneity of the second discrete choice, thus providing support for possible starting point bias (32). This bias may be in one of several directions. First, once a respondent has ‘made a commitment’ by indicating a willingness to pay for the first amount they are more likely to also be willing to pay the second amount than they would had they not received the first offer. Second, respondents may interpret the first value as the average

---

5 See Herriges and Shogren (37) for an explanation of this.

6 It is quite possible that respondents have incomplete knowledge about their true valuation of a nonmarket resource and thus may give wrong yes/no answers on whether to pay or not. This logic is that if the respondent knew the value preference with certainty then they would have no difficulty in specifying it when facing an open-ended survey. However, given the experience that individuals face difficulty in doing this, it seems likely that there is some range within which the value lies, but as to the precise value the individual is uncertain (39). That is, the discrete format may help "respondents who are 'information poor' regarding their actual WTP. Even if they don't know the maximum they would pay, they may know if a given amount is acceptable, too high or too low" (41, p198). Following this, Li and Mattsson (62) extend the work of Hanemann (23) and Cameron (59) by introducing preference uncertainty with individual responses. However, this thesis applies equally to open-ended data, and so is not an issue here. It is, however, interesting to note that the authors recognise that the logic of their approach may be extended to the open-ended survey approach.
social value, or price, of the good and baulk at being asked to pay ‘more than it costs’. Third, if respondents are not willing to pay the first amount they may feel guilty about it and be more willing to pay the second, lower amount, than they otherwise would. Finally, if they had said no to the first bid, they may become annoyed at being offered a lower bid, perceiving the interviewer to be trying to eke at least some money from them by lowering the bid, that they say no to the second bid even though they would have been likely to say yes had it been the only bid. One recent study (32) concluded that the most dominant effect on the mean, and distribution of, WTP was that respondents who initially said yes where inclined to persist in saying yes, and those who initially say no are likely to persist in saying no. ie starting point bias.

The second first-response bias of importance is ‘yea-saying’. This form of bias is, particularly, of potential importance in discrete choice surveys (21, 40). There is evidence to suggest that open-ended and discrete choice surveys do not demonstrate convergent validity (2, 42), and that this may be due to ‘yea-saying’, whereby respondents wish to agree with the interviewer regardless of their ‘true’ WTP. For example, the authors of one study comparing a payment card open-ended questionnaire and a single bound discrete questionnaire found that ‘yea saying’ had a positive influence on the results of the discrete responses compared to payment card approach (40). These authors conclude that “the results of this study strongly reject the hypothesis that dichotomous choice and payment card methods for the contingent valuation of a public good are procedurally invariant” (40, p130). Another study, which specifically tested for issues of optimal bid design and potential bias in discrete choice (single and double bounded) surveys for CV, illustrated that ‘yea saying’ was a particular problem that might be encountered, with 20% of responses estimated to be ‘yea-sayers’ (28). For convergent validity, neither procedure is assumed to be a truer measure of the construct than the other. However, such studies are important as they empirically identify sources of bias in discrete surveys which might be hypothesised to be insignificant, such as ‘yea-saying’.

A final point to be made is that even if starting point, or yea-saying, bias does exist in any CV survey, open-ended or discrete, this may simply reflect a need by the respondents for some context, or ‘bounds’ to the question (6). If this is true, then we should not, necessarily, be looking to remove such bias, but finding means to assist the respondent in their valuation struggle. For example, by altering the ‘context’ of analyses to provide greater, or more detailed, information, such as suggesting the ‘cost’ of the intervention.

In summary, it is clear that the potential for some form of first-response bias may exist. However, it is also clear that this applies equally for discrete-surveys (particularly double/triple bounded) as for open-ended. Of particular importance may be the recognition that such bias may be indicative of a respondents need for greater information to enable the valuation search to be expedited, and more research should be undertaken to assess the importance of this.
3.4 Bid Vector to be Used in Discrete Surveys

An important disadvantage of the discrete survey, compared to open-ended, is that it requires a sampling plan, or bid vector, to be determined, plus a corresponding vector of sample sizes corresponding to each bid. Once the data is collected the qualitative dependent variable data must be econometrically analysed and an expression for WTP (or WTA) chosen, which is a function of parameter estimates. There is little consensus over how to deal with either of these issues, which is problematic as the WTP value can differ markedly depending on how both these steps are performed (63, 64). In this section, the paper considers the evidence concerning the number and placing of bids, with statistical analysis issues covered in the next section.

Of critical importance for a discrete survey is that a bid vector, or set of bids to be offered to respondents, has to be chosen. That is, in order to offer bids for acceptance or rejection, one needs to estimate what the appropriate bids are. The correct number, and size of bid, is vital if calculation of the ‘true’ mean/median WTP is to be performed accurately. For example if a large proportion of respondents accepted the highest bid presented then that may indicate that the highest bid was set too low, thus biasing downward the resultant WTP value estimated. In order to establish this range, or vector, of bids, a pilot study is generally required, which uses an open-ended survey to establish relevant bids (or the most likely range of bids and likely mean) and likely proportion of positive responses to them (a refinement of this may be to perform the discrete survey in sections, revising the bids if necessary during the survey (1)).

The efficiency, and therefore cost, of the discrete survey also depends on the number of bids used, with judicious selection of bids reducing dramatically the typically large sample sizes required. Although recent reports suggest that, in theory, the most efficient number of bids is two (28, 65, 66), others have suggested a greater number (67, 68), and indeed more bids are generally used. This is both in order to test the adequacy of the distributional assumption for WTP (or log WTP) associated with the fitted binary response model, and also because the theoretically efficient models require more information concerning the true WTP than is available in practice when designing such surveys. However, although the importance of the bid vector is undeniable, the importance of the actual number of bids is questionable. For example, there are those who argue that the length of bid vector, and number of bids in the tails of the distribution have a large impact on estimated mean WTP (69), whereas others have argued that more, or less, complete coverage of the range has only minimal impact on the biases of the WTP estimate, and so placing bids in the ‘tails’ essentially wastes them (28). For example, Cooper and Loomis (69, 70) focus on the variation caused by changes in the number of bids. They estimate, by analysing data as bids are successively removed, that the mean WTP changed from between a reduction of 63% (removal of four lowest bids) to an addition of 37% (removal of every other bid). Variation also differs according to model, with removal of lower bids more important than

It appears to be a moot point that, in effect, the results of the (‘valid’) discrete survey rely upon the response given in a smaller (‘invalid’) open-ended survey!
higher bid removal in some models. The specification of bids to be used is therefore critical to the analysis.\(^8\)

In contrast, using Monte Carlo simulation methods, Alberini (29) found no increase in power when greater than 6-10 bids are used, or when bids are placed far out in the expected tails (where the probability of a yes response is less than 2-3%). Although double bound questionnaires have greater power, even here there is no power advantage in having the number of initial bids greater than 4. It was found that with sample sizes of even 1,000 the power to detect goodness-of-fit of the model is as low as 0.3. That is, a test of adequacy of the assumed model (e.g., Pearson chi square) will reject the null hypothesis only 30% of the time - thus leading the researcher in most cases to conclude that their model is reasonable when in reality the true WTP is very different to the assumed one. The conclusion of this paper was that single bound discrete choice questions tend to have little power in detecting incorrect distributional assumptions for sample sizes commonly used in CV surveys (n<1,200), and only become powerful with sample sizes greater than 2,400. This is undesirable, to say the least, since the mean and median WTP can vary greatly with the assumed distributional model, thus greatly influencing the resultant CBA (see section 3.5 below for more detailed discussion of these issues). In contrast, double bound questionnaires, with up to 4 initial bids, give greater power (0.9 at n = 1,000), and the author concludes that double bound questionnaires are therefore preferable, with no more than 4 initial bids which should not be placed in the extreme tails of the expected distribution (as a general rule of thumb, the suggestion is not to place bids outside the 15\(^{th}\) and 85\(^{th}\) percentiles for single bounded, and the 10\(^{th}\) and 90\(^{th}\) percentiles for double-bounded surveys).

A final point to note is that, in addition to the number and placing of bids, the estimate of mean/median WTP can depend upon how the total sample is distributed between a given bid amount, and placing of bids (72), and the “sensitivity of closed-ended valuation results to sample selection is an issue that merits further consideration” (60, p305).

### 3.5 Statistical Analyses Required to Calculate Mean WTP

The use of discrete data requires statistical manipulation. However, the ‘correct’ statistical approach is by no means clear, and there remain several issues, such as the correct model specification and functional form, estimation of confidence intervals and the issue referred to above of the ‘tails’ of the distribution. This section briefly discusses issues of model specification and confidence intervals. The critical point, however, is that the mean (and to a lesser extent the median) has been shown to vary considerably with the model’s functional form and assumptions concerning the ‘tails’ of distributions.

---

\(^8\) Although it should be noted that it has been argued that the results of this study are due to the wrong model being fitted to the data (high variance around mean WTP value), and that as long as the correct model is fitted, the only important bids are those close to the true mean (infact, bids in the tails may be counterproductive) (71).
3.5.1 Model Specification and Functional Form

Model specification essentially depends on four factors: (i) distribution assumed for WTP (e.g., normal, log normal or Weibull); (ii) the protocol for eliciting WTP (e.g., double versus single bound); (iii) constraints imposed on design (such as number of bids, values of bids, sample size); and (iv) the WTP statistic of interest, such as median or mean. Although these are all of importance, as discussed elsewhere in the paper, perhaps the largest impact on the resultant estimate of value is the statistical design of models for estimating the WTP measure (mean or median) (46), and so an overview of designs argued to maximise the precision of the statistic of interest is useful.

WTP as estimated using discrete data is sensitive towards the functional form with respect to bid and probability of acceptance (1, 46). Often some form of non-parametric test is used, which does not impose a functional form, and makes assumptions used in generating the mean WTP more explicit. However, using this method it is not possible to calculate the effect of different explanatory variables on WTP (e.g., income), and so in practice often the latter, or both, are used in analysis. When considering the statistical analyses the choice is essentially between some form of parametric, nonparametric or bivariate models. These models tend to apply to both single and double bounded surveys, although multilevel modeling has been proposed to be used with triple bounded data (73).

Alberini (24) discusses fully parametric models which she believes are the optimal designs for discrete choice contingent valuation surveys, if the measure of interest is the median. In analysing both single and double-bound surveys, the MLE method is often used (74), based on a log likelihood function, using a logarithmic transformation of WTP as a log normal variable. With such a function an asymmetric distribution means that the mean and median will not coincide (which of course has implications for which is therefore used, and is discussed in section 3.6). There are various measures by which the efficiency of estimation of the desired statistic may be improved, such as minimum variance, fiducial and d-optimal designs. However, the main point made is that “designs which are good for one measure of welfare, such as median [WTP], tend to perform worse for other measures of welfare (mean [WTP]), if [WTP] is log normal. This result is likely to hold for other asymmetric distribution commonly used with CV survey data, such as the log logistic and the Weibull” (24, p297). It is interesting to note that Alberini makes the point that if the values in a double bound survey are not from the same underlying true value, as mentioned earlier, then an additional parameter of interest is the correlation coefficient associated with these two values, which requires the use of a bivariate probit model (24). However, overall there appears to be a trade-off between designs which do well with respect to estimating median WTP, and those which do well with respect to mean WTP. There is no design which will be efficient with respect to both mean and median WTP simultaneously, the implications of which are further discussed in section 3.6.

Kanninen (66) also reports on the optimal experimental design for discrete surveys, although concentrating on double-bounded dichotomous choice contingent valuation, and in particular addresses the optimal choice of bid values for double bound logit model, derived to satisfy d-optimal, c-optimal and fiducial efficiency criteria. Kanninen notes that sequential experimentation,
where results are continually monitored and bids continually updated on the basis of this information (although this is likely to be extremely difficult with mail out surveys, it would be quite feasible with interviewer (face to face or phone) administered survey), is likely to improve the efficiency of estimation of WTP (75). Of importance is the conclusion which Kanninen draws that “it would be useful to consider the potential gains of a triple bound approach, where a third follow up bid value would narrow the individual WTP range even further. Such a procedure would surely obtain more efficient results than the double bounded procedure, since more information per observation is better than less. Clearly, the triple bounded model would provide fairly small gains over the double bounded model in terms of the c-optimal criterion, as the gains are fairly small for the double bounded model over the single bounded. It appears likely that for the d-optimal criterion, the triple bounded procedure could indeed offer significant gains. In addition, the triple bounded procedure would be further insurance against misspecification of the parameter values, and a poor bid design” (66, p145). It seems as if Kanninen is providing a powerful argument for use of the bidding system employed in open-ended surveys, if we substitute ‘triple bounded’ in this quote for ‘open ended bidding’!

However, the main problem with parametric approaches of any kind is that an assumption must be made concerning the likely distribution of values. As such an assumption is required, one faces the risk of misspecifying the functional form of the model. Model misspecification (such as through omitted variables, heteroskedasticity or distributional asymmetry) presents problems because it leads to inconsistent parameter estimates (76), and in the case of WTP leads to erroneous welfare measures (77). Not only may it be argued that such misspecification is highly likely, since one is required to make an assumption about something which cannot possibly be observed, it has been empirically shown to effect the welfare measure used (77). There have therefore been various non-parametric approaches to the estimation of WTP from discrete choice contingent valuation studies proposed (31). The non-parametric approach is proposed because it does not require any assumption with respect to distribution, plus the simplicity of such an approach, compared to parametric analyses, means that computations can be done “on the back of an envelope” (31, p135).

Perhaps the most interesting recent development is the prospect of using multilevel modelling to analyse discrete choice data (73). Although this has been used primarily to analyse triple bound data, it is interesting to note that the extension of the multilevel modelling approach to n-bounded data is a trivial issue (73). Briefly, data from a triple (or n) bounded survey can be seen as hierarchical in nature, with each individual being measured repeatedly by yes/no responses to a maximum of three (n) separate questions, and therefore being three (or n) level in nature. Such a multilevel approach has been theoretically outlined and empirically applied (73, 78), and for more detail concerning this method the reader is referred to Goldstein (79). The important conclusions of such research are that “rather than sampling many individuals at relatively few bid levels to ensure reliable estimates for each sample at each bid level, it would be more efficient (provide more useful information) in the multilevel model to have fewer individuals and many bids……Similarly, having a bidding game with more questions, or bounds, would simply increase the amount of useful information available with a minimal increase in complexity of the model” (73, p206). The authors are currently researching into this (80). This is important as it seems to
be bringing us full circle back to open-ended bidding surveys, although with a more advanced form of statistical analysis.

Further argument for this is provided by the authors in a comparison of an open-ended survey mean WTP with the mean taken from the triple bound survey had it been a double bounded survey (ie the mean with the information from the last bid excluded) (73). Such comparison revealed that the mean from the discrete choice survey was more than twice the open-ended mean. However, when the triple bound estimates were analysed, the mean was reduced to a level only slightly above that of the open-ended mean. From this data the authors conclude that “there has been much debate on the possible inflation of discrete choice WTP estimates compared to the results of open-ended questions, but here we can see that, with the extra information provided by the triple bound discrete choice analysis, the estimate of E(WTP) has fallen to be comparable with the open-ended results” (73, p208). That is, the information from open-ended surveys is valuable, and that triple bounding is a move toward n-bounding, in which case we are back at the bidding approach by a different route. The authors also show that in their study starting point bias did not seem to exist. If it does not for triple, then likely would not if they had continued to n-bounded. Perhaps multilevel modelling is opening up the analysis of open ended data such that this technique will again be favoured, as the authors seem to indicate themselves.

3.5.2 Confidence Intervals

Another issue of importance, apart from the functional form of the model chosen to analyse WTP data, is how reliable the estimated WTP value (mean or median) is from a statistical point of view. i.e., the variance of WTP which may be represented by placing confidence intervals around the mean at a specified level of significance. Confidence intervals have been estimated a number of ways, including the delta method (81), bootstrapping (82) and Monte Carlo simulations (83, 84). For example, Cooper (85) provides a comparative analysis of the reliability and precision of four techniques for estimating confidence intervals: the bootstrap (86), the jackknife (87), and those proposed by Krinsky and Robb (88, 89) and Cameron (90). Using Monte Carlo analysis, Cooper concludes that although all four measures perform well on average, they differ in the frequency with which they do well, and that the best choice depends on the sample size, distribution of the WTP estimate and choice of functional form for estimating WTP. e.g., bootstrap being the preferred choice where sample size is over 1,000, but the Cameron method is preferred where the sample size is less than 100 (provided the distribution is logistic; the Krinsky and Robb performs best with Weibull). The results of this study again highlight the complexities of estimating precise WTP estimates from discrete data. Similarly, Bowker and Stoll (91) also indicate that “models with fairly similar statistical fits can lead to very disparate measures of economic value, regardless of whether the mean or median is chosen” (91, p379).

However, such calculations as bootstrapping and jackknifing, tend to use parameter estimates, and their standard errors, from multiple regression equations. It is rare that the overall goodness-of-fit of the equation is included, and so potential bias from poorly fitting models can result. One approach to address this, by including unexplained variance explicitly, is generalised linear mixed
model (GLMM) regression (92), which includes both fixed and random terms in the model fitting process. Such analysis has been used to illustrate that modelling unknown, random, elements can provide significantly different results from modelling fixed, known, effects only (92). Such a result presents a serious concern, and suggests that modelling of these uncertain variables be included in statistical analysis of discrete choice data.

A final point is that it has been suggested that the cumulative distribution functions (cdf) estimated from discrete data may have ‘unrealistically fat’ right hand tails, causing mean WTP to be overestimated. Fat tails are caused either by the study not including bid levels higher than most respondents WTP, or from ‘yea-saying’. Truncating the range of integration will result in a lower bound estimate of true mean WTP (63, 93), and a normalisation procedure may depress the estimate further (94). An alternative procedure has been outlined (‘pinched logit’, whereby the cdf is assumed to reach 1 at maximum bid, but doing so without a discrete jump), although it entails extrapolation beyond the range of data and as such tends to be highly variable (95) (in effect this approach is similar to the non parametric approach of Kristrom (31)).

In summary, we can see that there is a great complexity of statistical analysis brought about by the use of discrete choice survey designs, as compared to open-ended questions. No one model has yet to emerge as ‘dominant’, and analyses indicate that the resultant WTP value can vary widely based on the model specification and functional form. Although no definitive conclusion can be gained from this concerning the ‘validity’ of WTP estimates using discrete versus open-ended designs, it seems likely that the potential for not estimating ‘true’ preferences is just as great using discrete data, if not greater, given such analytical issues, as it is with open-ended surveys.

### 3.6 Use of Median Versus Mean Estimate of WTP

When using logistic regression to analyse discrete data, the mean WTP is highly sensitive to assumptions concerning the functional form (42). In contrast, the median is relatively stable (1, 23)\(^9\). As these measures often do not coincide (in fact rarely) in the statistical analyses used to analyse discrete data, the issue becomes which measure to use. Such an issue has been recognised since Hanemanns seminal paper in 1984 (23). Hanemann, for example, acknowledges that the mean will be sensitive to slight changes in the shape of the distribution resulting from different estimation methods or outliers, whilst the median is relatively robust. He also acknowledges that the choice between them entails a value judgement as to the appropriate method of conducting welfare evaluations. This has led to a (renewed) debate in the literature concerning the most appropriate summary statistic to use as a point estimate of WTP (23, 63, 64).

It has been argued that there is only one theoretically correct measure; the mean (48, 63). The ultimate goal of such analyses is to compare the projects expected aggregate costs and benefits.

---

\(^9\) Although it has been reported, in directly comparative studies of discrete versus open surveys, that there is no significant difference in mean or median values (39).
That is, compare the total WTP for the project against the projects total costs (leaving equity and income distributions aside for the moment). The mean is the correct statistic which, when multiplied by the number of persons in the population who receive the program under evaluation, yields the total population (societal) WTP (1). If, for example, there are 100 individuals concerned, then total WTP is simply 100 times the mean value. The mean is the natural, and theoretically valid, measure to use. However, often the results of discrete surveys are interpreted as a referendum, where the median is the relevant value used (that point at which 50% of the sample are in favour, and 50% against, the proposed project at a cost represented by the median). The problem with the median voter approach is that it is known to not, in general, produce a Pareto-efficient outcome (63), as median voters prefer more/less public expenditure than is consistent with Pareto efficiency. Two points arise from this. First, the median appears to be a pedagogical device which may help the respondent understand the intuition behind the WTP questionnaire. However, it is not the correct statistic to use in calculation of cost-benefit. Second, if one is simply interested in the outcome of a referendum then one could simply ask ‘voters’ if they would plan to vote yes or no. It would be unnecessary to perform a full CV experiment concerning actual WTP (63). What we therefore witness is the use of a theoretically incorrect measure because of its statistical advantages. Although a counter argument is that the mean only differs from the median in such studies due to an artefact of the CV study itself skewing the WTP data due to a few individuals being WTP very high amounts. Thus using the median could be argued to be closer approximating the true mean WTP.

However, much of the argument becomes ethical, based on whether the criterion of Pareto efficiency is one which is considered the most valid in the decision of whether a project should proceed or not. For example, Hanemann (64) advances the example that “out of a community of 1,000 individuals, 999 value some environmental amenity at $1 each, while one person values it at $1,000. If there were some project to protect this amenity at an average cost of $1.98 per person, and if the cost were to be shared equally among all the members of the community, I would regard it as a bad project - even though it passed the Kaldor-Hicks test” (64, p1060). Hanemann recognises that his position would disenfranchise those who have the largest stake, but suggests that this may occur every time a majority rule is taken. Perhaps then the issue rests on more ethical grounds of income distribution, for example in assuming that one individual in Hanemanns example can afford to pay $1,000, whereas the other cannot. If this is not the case, that for example income distribution is equal, then the case for not accepting the Pareto efficient position seems hard to defend, as the issue in interpersonal comparison of utility is that the person representing $1,000 values the good 1,000 times the others. Let me phrase this another way - suppose the good in question is one persons life - person A. Now let us run the example with that person WTP $1,000, but the others only WTP $1 to save his life. What would be the optimal criterion now? I would suggest Pareto efficiency. However, this argument shall not be considered further here, and will be left for discussion elsewhere.
4 DISCUSSION

It is clear from the review presented in this paper that the debate concerning the appropriate use of discrete, as compared to open-ended, techniques for valuing benefits in contingent valuation surveys is not as clear cut as it would at first appear. There are issues yet to be resolved concerning the degree of bias within the open-ended technique itself, but more importantly there appear to be substantial additional issues with the use of discrete survey techniques, whether single, double or triple bounded. While the discrete choice questionnaire seems to have been favoured, major problems still remain in its implementation. These problems arise from the fact that, simply, discrete choice survey design and analysis is very complex. Survey design elements include total sample size, bid range, specific bid levels, allocation of the total sample among the bid levels, and form of statistical model used to analyse such data. None of these issues has by any means been clearly resolved.

The basis for the use of discrete surveys in preference to open-ended seems to rest on some combination of: (i) a belief that this technique offers a more ‘realistic’ market, and will therefore lead to more valid responses (a truer estimate of actual WTP by respondents); and (ii) the supposed tendency of discrete surveys to yield higher response rates, through reduced mental demands (especially for mailed surveys). Both of these issues have been addressed, and conclusions disputed, in this paper, along with some others of importance. It is undoubtedly the case that open-ended WTP surveys contain issues which need consideration, and the technique is far from being fully developed. However, to attempt to solve, or perhaps more appropriately avoid, these issues by adopting a very limited version of the traditional bidding game (using only 1, 2 or 3 bids) seems foolhardy. Not only might such a technique still present problems of survey cost and first-response bias, but it also brings unique problems of its own, such as issues concerning bid vectors and appropriate statistical analyses. Furthermore, although estimates from open-ended and discrete surveys have been shown to provide markedly different WTP results, there is no conclusive theoretical or empirical justification for one being considered more valid than the other.

It is particularly puzzling why the discrete technique has moved toward double or triple bounding. It is clear that there is some support for the use of these methods in obtaining more precise estimates of WTP and reducing the sample size required by single bounded response. However, it would appear that to progress beyond a single bound creates a bidding game approach, with the number of bids simply truncated early on in the game. The logic that argues for double or triple bounding would seem, to this author at least, to apply equally well to n-bounding, which then brings us back to the traditional bidding game. One might further argue that even the single bounded discrete choice survey design is just a bidding game with one bid. It would therefore seem to this author that there is some agreement that some form of bidding is the way to go. The question thus becomes one of how many bids, the relative benefits from using only one compared to more than one, and then the benefits of two versus three versus four versus n.

This author would therefore suggest that the debate over discrete versus open-ended techniques is a red-herring in the search for ‘valid’ WTP values derived from CV surveys. There appears no
reason why the open-ended survey should be summarily dismissed, and research may better be targeted to refinement of the bidding approach to reduce bias. For example, perhaps the presence of starting point bias indicates that respondents simply cannot find any unique, precise, WTP value, but require some assistance in finding a focus, or bounds, within which to provide a range of values they would be prepared to pay. That is, it is quite possible that respondents have either incomplete knowledge about their true valuation of a nonmarket resource, or do not possess a unique ‘point estimate’ valuation. In either case there is therefore a range of values which they would be prepared to pay, with a corresponding probability distribution. It is the possibility of this issue causing bias that should therefore be addressed.

Perhaps as a final point, it is worth noting that there has been (to this authors knowledge) no attempt to calibrate data from discrete surveys (or open-ended in the case of health care) against real decisions, and so ultimately we cannot establish which is the most ‘valid’. This is important to bear in mind when considering the current recommendations of most practitioners of WTP that discrete surveys are to be preferred as they are argued to be more ‘valid’. i.e that they may approximate market decisions more closely and hence provide more realistic data. To resolve the issue there clearly needs to be research which compares both type of survey with real financial transactions. Until such studies are undertaken we are left with no definitive answer as to which is more ‘valid’, only a series of opinions from which the researcher must ultimately draw their own conclusion. As another view in this debate, this author would recommend caution in the seemingly indiscreet adoption of the discrete choice question approach at the present time.


8 Donaldson C, Thomas R, Torgerson DJ 1995, Open-ended versus payment scale approaches to eliciting willingness to pay, Discussion Paper 01/95, Health Economics
Research Unit, Departments of Public Health and Economics, University of Aberdeen, Aberdeen, Scotland.


