ELSEVIER

The contingent valuation method: a review

L. Venkatachalam*

Institute for Social and Economic Change, Nagarbhavi, Bangalore 560 072, India

Abstract

The contingent valuation method (CVM) is a simple, flexible nonmarket valuation method that is widely used in cost–benefit analysis and environmental impact assessment. However, this method is subject to severe criticism. The criticism revolves mainly around two aspects, namely, the validity and the reliability of the results, and the effects of various biases and errors. The major objective of this paper is to review the recent developments on measures to address the validity and reliability issues arising out of different kinds of biases/ errors and other related empirical and methodological issues concerning contingent valuation method.
© 2003 Elsevier Inc. All rights reserved.

Keywords: Contingent valuation method; Willingness to pay; Embedding; NOAA

1. Introduction

The CV method is a widely used nonmarket valuation method especially in the areas of environmental cost–benefit analysis and environmental impact assessment[1] (see Mitchell and Carson, 1989; Cummings et al., 1986). Its application in

---

* Tel.: +91-80-321-5468/5519; fax: +91-80-321-7008.

E-mail address: venkat@isec.ac.in (L. Venkatachalam).

[1] The other areas in economics where the CV method is increasingly being applied are: health economics (e.g. Thompson et al., 1984; Johannesson and Jonsson, 1991; Johannesson et al., 1996), transportation safety (e.g. Jones-Lee et al., 1995) and cultural economics (Hansen, 1997; Thompson et al., 2002). Also, the CV method is currently being integrated and interlinked with other 'stated preference methods' such as 'conjoint analysis', 'stated choice analysis' and 'choice experiments' (see, Louviere, 1988; Louviere et al., 2000; Louviere and Woodworth, 1983). For critical review of stated preference methods, please see Kristrom and Laitila (forthcoming). Since reviewing studies on these topics in detail may not be possible within the scope of this review, the focus of the present review is restricted mainly to CV studies in environmental economics.

environmental economics includes estimation of *non-use values* (e.g. Walsh et al., 1984; Brookshire et al., 1983), *nonmarket use values* (e.g. Choe et al., 1996; Loomis and duVair, 1993) or *both* (e.g. Niklitschek and Leon, 1996; Desvousges et al., 1993) of environmental resources. In recent years, this method is commonly used in developing countries to elicit the individuals' preferences for the basic infrastructural projects such as water supply and sanitation (see Whittington, 1998; Merrett, 2002). Though a popular nonmarket valuation method, a group of academicians criticise this method severely for not being a proper method of estimating the nonmarket values (see Hausman, 1993). Hence, the major objective of the present paper is to review the developments and issues on the theoretical, methodological and empirical aspects of CV method.

The CV method was originally proposed by Ciriacy-Wantrup (1947) who was of the opinion that the prevention of soil erosion generates some 'extra market benefits' that are public goods in nature, and therefore, one possible way of estimating these benefits is to elicit the individuals' willingness to pay for these benefits through a survey method (see Portney, 1994; Hanemann, 1994). However, Davis (1963) was the first to use the CV method empirically when he estimated the benefits of goose hunting through a survey among the goose-hunters. This method gained popularity after the two major non-use values, namely, option and existence values,[2] have been recognised as important components of the total economic values in environmental economics literature, especially during the 1960s. While the conventional revealed preference methods such as travel cost method are not capable of capturing these non-use values (Smith, 1993), the only method that is identified for estimating these values is the contingent valuation method (CVM) (see, Desvousges et al., 1993). Hence, a considerable amount of studies on CVM—both theoretical and empirical in nature—have emerged in the economic valuation literature, including a large number of studies criticising the CV method.

2. Validity and reliability of CV studies

Summaries of CV studies[3] by different authors reveal that the major criticism of results of CVM revolves mainly around two aspects, namely, (a) validity and (b) reliability (Smith, 1993; Freeman, 1993; NOAA, 1993). In simple terms, validity refers to the 'accuracy' and reliability refers to 'consistency' or 'reproducibility' of the CV results (Kealy et al., 1990). In other words, validity refers to the degree to which the CV method measures the theoretical construct of

---

[2] While option value refers to the individuals' willingness to pay for preserving a particular resource base for future use, the existence value refers to the individuals' willingness to pay for preserving a particular resource merely for the sake of its existence (see Freeman, 1993).

[3] See Smith (1993) for a good summary of CV studies.

interest which is the true economic value[4] of individuals (Freeman, 1993). The validity is of three types, namely, content validity, criterion validity and construct validity (Mitchell and Carson, 1989; Bateman et al., 2002). The content validity in a CV experiment simply refers to the ability of the instruments included in the scenario to measure the value in an appropriate manner. Criterion validity of the CV method may be assessed in terms of another measure, say a 'market price' for the same commodity which may be considered as a criterion. Construct validity has two forms: convergent validity and theoretical validity. The convergent validity refers to the correspondence between two measures of the same theoretical construct. CV results can be said to be 'theoretically valid' if the results conform to the underlying principles of economic theory. In other words, the theoretical validity involves assessing the willingness to pay (WTP) values of the CV method by way of regressing the WTP value against standard economic variables (Mitchell and Carson, 1989). On the other hand, the reliability refers to extent to which the variance of the WTP amounts is due to random sources (Mitchell and Carson, 1989). According to Loomis (1990), 'Reliability requires that, in repeated measurements, (a) if the true value of the phenomenon has not changed a reliable method should result in the same measurement (given the method's accuracy) and (b) if the true value has changed a reliable method's measurement of it should change accordingly' (p. 79). The following section is devoted to discuss various errors/biases affecting the validity and reliability of the CV method along with illustrations from empirical studies that looked into addressing these two issues.

2.1. Disparity between WTP and WTA (willingness to accept)

Hicks (1946) classified the 'consumer surplus measure' into two different categories, namely, the compensating variation and the equivalent variation. For a 'proposed welfare gain' due to provision of public good, the compensating variation refers to the amount of money income that has to be given up by the consumer to attain increased level of utility (i.e. WTP measure).[5] The equivalent variation refers to the amount of compensation required to be provided to the individual so that she could attain an improved utility level in case the provision of the public good does not take place (i.e. WTA). For a welfare loss, the compensating variation refers to the amount of money income that is required to compensate the individual for the welfare loss experienced (i.e. WTA) and equivalent variation refers to the amount of money income to be sacrificed by

---

[4] The true economic value refers to the Hicksian consumer surplus measure—either the equivalent variation or the compensating variation. Alternatively, the 'true measure' could be described as the one which would have existed if a real, perfectly competitive market had existed for the good under consideration (see Mitchell and Carson, 1989).

[5] From the point of view of the policy-maker, it is the money income that needs to be taken away from the consumer, which will place her on the same indifference curve prevailing before the provision of the public good.

the consumer to prevent the loss from occurring in future (i.e. WTP) (Bateman and Turner, 1993).

In rigorous economic terms, the contingent valuation method estimates the Hicksian consumer surplus—either the compensating variation or the equivalent variation—due to change in the provision of public goods (Bateman and Turner, 1993). One of the important conclusions derived from the above paragraph is that, in principle, either WTP measure or WTA measure could be used interchangeably to elicit individuals' preferences for change in the level of environmental goods and services.[6] Yet, one of the issues that is supposed to affect the validity of the CV results is the disparity that arises between the WTP value and WTA value for the same good under consideration (Mitchell and Carson, 1989). It has been demonstrated both theoretically as well as empirically that the WTA value is always greater than the WTP value if used for the same issue (Shogren et al., 1994; Hanemann, 1991; Brookshire and Coursey, 1987; Coursey et al., 1987; Knestch and Sinden, 1984; Bishop and Heberlein, 1979; Willig, 1976). This being the case, the question that has to be addressed is: which measure (i.e. either WTP or WTA) should be used in a CV survey to elicit the value of the changes in the provision of public goods? (Mitchell and Carson, 1989). Before going on to answer this question, let us see the reasons why the disparity between WTP and WTA measures arises.

2.1.1. Reasons for disparity between WTP and WTA

The WTP/WTA disparity has been attributed to many different factors. Willig (1976), having theoretically demonstrated the disparity between WTP and WTA measures for a priced commodity, concluded that the WTP/WTA disparity could be attributed to the 'income effect'. In a strict sense, the income effect in economics refers to the effect of additional income on the quantity purchased of a particular commodity (Diamond et al., 1993). The underlying implication of the income effect is that the WTP for a good is constringed by income, whereas the WTA compensation is not. In this sense, Willig's (1976) conclusion may be interpreted as to imply that WTP and WTA would diverge for a commodity which has high income elasticity of demand[7], while accompanied by income constraint. Randall and Stoll (1980), extending Willig's (1976) analysis from the 'price space' to 'commodity space', proved that the WTP and WTA measures are found

---

[6] Take for instance, a proposed change in an environmental policy that would result in an improved air quality in a particular locality. *Ceteris paribus* (such as property rights, etc.), an individual in the locality can either be asked to state her maximum willingness to pay (compensating variation) for ensuring the change in the policy that aims at say, improving the air quality in the region or she can be asked to state minimum willingness to accept compensation (equivalent variation) required to compensate the expected utility foregone due to nonimplementation of the proposed policy.

[7] However, Carson et al. (2001) argue that it is not the *income elasticity of demand* but the *income elasticity of WTP* (i.e. change in the WTP for a fixed quantity, as income changes), which is more relevant in estimating the value of public goods.

to be closer for public goods unless they are affected by an unusual income effect (Hanemann, 1991). In explaining the disparity further, Hanemann (1991) showed that: (a) for a change in the quantity (unlike change in the price), there is no such presumption that WTP and WTA measures should be close; and (b) not only the income effect but also the 'substitution effect'[8] explains the larger disparity between WTP and WTA. It means that the divergence between WTP and WTA values could range from zero to infinity depending on the degree of substitution between goods, accompanied by positive income elasticity (Shogren et al., 1994).

Some of the empirical studies have attempted to understand the influence of substitutes on the WTA/WTP disparity. For example, using two private goods (a candy bar and a coffee mug) that are close substitutes and a public good (reduction of human health risk), which has no close substitute, Shogren et al. (1994) demonstrated that the divergence between WTP and WTA values for the two private goods disappears after repeated trials, while the divergence is robust and persistent for the public good. The results of this study not only proved the role played by the substitution effect in explaining the disparity between WTP and WTA but also showed how the respondents' 'familiarity' with the valuation experiment plays a role in reducing the disparity. In another empirical attempt, Adamowicz et al. (1993) used two goods—one without substitute (a feature film in a local theatre) and the other with substitute (hockey match with a substitute of live radio/TV telecast). Their experiment demonstrated that substitution factor does explain the disparity between WTP and WTA estimates. However, the authors argue that even though the substitution factor reduces the disparity it does not eliminate the disparity entirely. This implies that there are other factors that could provide possible answers to the WTA/WTP disparity.

Apart from income and substitution effects, some of the theoretical developments in economics and psychology are also extended to explain the WTA/WTP disparity. For example, the 'prospect theory'[9] developed by Kahneman and Tversky (1979) also explains the variation in the WTP and WTA measures. According to Kahneman and Tversky (1979), the loss of a commodity for an individual is considered to be greater than the gain derived from buying the same commodity. This is because, increase in income is weighted by a relatively small utility compared to decrease in income which is weighted by much larger utility (Coursey et al., 1987). More precisely, for the 'loss averse individuals' the disparity would be greater (see Brookshire and Coursey, 1987). Thaler (1980) demonstrated WTA/WTP disparity through 'endowment effect' and noted that it might be related to prospect theory, which is about decision making under uncertainty. The existence of loss aversion by individuals has been proved by Brookshire and Coursey (1987) who found in a hypothetical experiment that the respondents'

---

[8] The substitution effect simply refers to impact of the availability of substitute commodities on the quantity demanded of the good in question.

[9] Prospect theory refers to the asymmetric valuation of goods on the claim that selling a particular good generates loss and buying the same good generates gain.

WTA value for a particular level of *reduction* in an area with tree cover was 75 times greater than the WTP value for the same level of *expansion* in the same area.

The concept of 'property rights' as well as 'transaction cost' after Coase (1960) also holds the reason for the disparity between WTP and WTA. For instance, the WTA/WTP disparity has been attributed to the rejection of the underlying property rights by individuals (Mitchell and Carson, 1989). According to Mitchell and Carson (1989), in many cases the individuals consider the WTA property right as illegitimate or implausible or both. This is evident from generation of large number of 'protest bids' in CV studies that used WTA measure as elicitation format. Apart from property rights, the transaction costs involved in the process of obtaining goods and services may also result in WTP/ WTA disparity (Brown and Gregory, 1999). During market transaction, the consumer calculates the total cost of buying and selling the good that includes not only the price of the good but also the transaction cost. This being the case, while buying the good the individual's willingness to pay would *exclude* the transaction cost and while selling the same good the willingness to accept compensation would *include* it. However, since market for public goods are either nil or very weak, the transaction cost theory may not be expected to play a major role in the WTP/WTA disparity issue.

Apart from theoretical aspects, the value elicitation or experimental aspects of CV method are also found to affect the WTA/WTP disparity. Cummings et al. (1986) bring the 'familiarity' issue (see also, Shogren et al., 1994) for addressing not only the disparity between WTP and WTA values but also many other biases in CV method. How the respondents' familiarity with the valuation experiment (as well as the good) and their optimization strategy in a CV study leads to provide answer to some of the serious problems in CVM have been empirically tested by many authors, especially in laboratory experiments (e.g. Coursey et al., 1987). These experiments show that repeated trials (that provide adequate time for the respondents to understand the issue and optimise) lead to convergence between WTP and WTA measures. In many cases, it was observed that WTA measure comes down gradually over repeated trials, whereas the WTP measure remains stable for number of trials (Coursey et al., 1987). These findings provide support for Hoen and Randall's (1987) argument that the disparity between WTP and WTA would be greater for those individuals who lack time to optimise their decision.

In general, studies on WTA/WTP disparity conclude that a larger difference between WTA and WTP can be attributed to the 'weak' experimental features of the CV studies. These features include hypothetical payments, students being used as objects, using less incentive compatible elicitation questions, etc. (see Horowitz and McConnell, 2002). The implication is that if the survey is designed 'as realistic as possible' then the WTA/WTP ratio would become smaller. However, results of a meta-analysis by Horowitz and McConnell (2002) provide different kinds of insights into the debate. Horowitz and McConnell's (2002) meta-analysis tested whether the high WTA/WTP ratios are an experimental artifact, or they truly present a broad-based picture of preferences. Forty-five

studies on WTA/WTP have been taken for the meta-analysis and these studies dealt with goods such as chocolates, pens, mugs, movie tickets, hunting licenses, visibility, nuclear waste repositories, nasty-tasting liquids, etc. Using a random effects regression model, the study regressed the WTA/WTP ratio against type of good (ordinary vs. otherwise), survey design (hypothetical or real payment; elicitation technique; and student and nonstudent), mean WTP and year. The results of this study are striking. For example, the WTA/WTP ratios in real experiments do not differ significantly from that of hypothetical experiments. This finding is contradictory to the conventional notion that the experiments involving real payment are superior to experiments involving hypothetical payments. Another important finding of this study is that WTA/WTP ratio is higher for incentive compatible elicitation techniques, whereas the earlier studies found that incentive compatible elicitation techniques provide low WTA/WTP ratio. Another finding which goes against the earlier findings is that the WTA/WTP ratio is significantly lower for students than the general public. This implies that moving the laboratory experiment to the real world situation will not reduce the ratio. Also, respondents' familiarity with the experiments does not provide lower ratios. All these important findings of Horowitz and McConnell (2002) suggest that the high observed WTA/WTP ratios do not appear to be experimental artifacts and indeed, the answer for this disparity comes from the broad-based preferences of the individuals.

From the above discussion, it is clear that there exists a disparity between WTP and WTA and this disparity is influenced by many different factors such as income effect, substitution effect, transaction costs, broad based preferences, etc. Moreover, the disparity has been an accepted phenomenon in the CVM literature as demonstrated by both theoretical as well as empirical studies. However, having agreed that there exists a difference between WTP and WTA, one important question that one needs to answer is: how much can it differ? According to Hanemann (1991), the WTA value can be five times greater than the WTP value. But some of the empirical studies show that the disparity between WTP and WTA for a same commodity ranges from a low of 2.4 times to a maximum of 61 times (e.g. Brookshire and Coursey, 1987). Occurrence of large-scale disparity between the WTP and WTA in empirical studies leads to the following conclusions: (i) one or both of the CV measures are wrong or the theory is wrong (Carson et al., 2001); and (ii) the WTA measure is not a proper measure of consumer surplus (see Adamowicz et al., 1993), and therefore, WTP measure rather than the WTA measure is the proper measure of value that should be used in the CV studies (Cummings et al., 1986; NOAA, 1993).

2.2. Embedding or scope effect

One of the major sources of error in the CV method is 'embedding' that is frequently reported in many CV studies (Bateman et al., 1997). Embedding refers to a phenomenon in which a wide range of variation is found to occur in WTP

value for the same good depending on whether the good is valued on its own or valued as a part of a more inclusive package (Kahneman and Knestch, 1992). The embedding is also called part–whole bias, disaggregation bias, sub-additivity effect, scope effect (see Cummings et al., 1986; Mitchell and Carson, 1989; Hanemann, 1994; Bateman et al., 1997) even though the meaning between these concepts differs (Boyle et al., 1994). The embedding effect is defined in different ways by different authors. According to Kahneman and Knestch (1992), the embedding occurs if 'the same good is assigned a lower value if WTP for it is inferred from WTP for a more inclusive good rather than if the particular good is evaluated on its own' (Kahneman and Knestch, 1992, p. 58). According to Harrison (1992), embedding occurs when the WTP value for one good differs 'insignifi-cantly' with the WTP value for a more inclusive good. The existence of the embedding effect is supposed to result in a situation where 'different surveys can obtain widely variable stated willingness-to-pay amounts for the same public good, with no straightforward way for selecting one particular method as the appropriate one' (Diamond and Hausman, 1994, p. 46). Embedding affects the validity of the CV results, and therefore, it is recommended that every CV study should have an in-built mechanism called, 'internal consistency test' to assess the validity of results (NOAA, 1993). One way of assessing the internal consistency of the results is to involve the results for 'scope test'. To understand the nature of embedding effect and how the scope test is used for assessing the validity of the results, let us discuss briefly about some of the studies that have looked into these issues.

Earlier CV studies have widely reported occurrence of embedding in their results. For example, results of a CV study by Kahneman (1986) on conservation of fish in lakes showed that WTP value for cleaning up all the lakes in Ontario (larger area) was only slightly higher than the WTP value for cleaning up lakes in one region (smaller area). Later in a famous study, Kahneman and Knestch (1992) used three goods, namely, 'improved rescue equipment personnel' (say, Good A), which is embedded in 'improved preparedness for disasters' (say, Good B), which in turn is embedded in 'environmental services' (say, Good C). These three goods were assigned to three independent sample groups in such a way that sample group 1 received Good A, sample group 2 received Good B and sample group 3 received Good C. It was observed that the mean WTP values provided by three sample groups for all the three goods separately were as follows: $135.91 for Good C; $151.60 for Good B and $122.64 for Good A. The difference between the mean values was found to be statistically insignificant, implying that the embedding effect has occurred in the results. Hence, Kahneman and Knestch (1992) conclude that the values elicited through CV are not true values but arising out of individuals' 'purchase of moral satisfaction'.

However, Smith (1992) claims that none of the conclusions of Kahneman and Knestch's (1992) study (KK study hereafter), which suggest that the embedding effect is a problem in CV results, are correct. To substantiate his claim, Smith (1992) provides three reasons: (i) the question asked in KK study to test the embedding effect was not conveyed in such a way that the respondents could

properly understand the intention of the researchers. More precisely, the reference and target levels of the provision of the inclusive good were not properly conveyed to the respondents; (b) the general implementation of the survey was found to have been improperly done. For instance, issues such as how the questionnaire was developed, whether it was pretested, how sample households have been selected, statistical tools used, etc., have not been properly addressed in the study; and (c) the interpretation of the results in line with economic theory was also found to be flawed. Therefore, Smith's (1992) arguments imply that the results of the KK study should not be used to assess the ability of the CV method to measure the nonmarket values.

Moving one step forward, Harrison (1992) not only criticises the way in which the goods were described in KK's study but also tries to explain the reason for the embedding effect with the help of 'Good Cause Dump Hypothesis' which refers to a situation where the respondents have a WTP value for a basket of good causes and dump the whole value for any good in the basket when asked for valuing a single good. Moreover, reanalysing KK's data with an alternative method, Harrison (1992) concludes that the KK study had employed inappropriate statistical and data handling procedures. However, Harrison's (1992) claim has been challenged by Nickerson (1993). Nickerson (1993) evaluates KK's version of hypothesis, data and statistical analysis and Harrison's version of KK's hypothesis, data and statistical analysis independently and concludes that it is Harrison's statistical and data-handling procedures that are inappropriate.

Some of the CV studies conducted subsequently did also report the occurrence of 'embedding' in their results. For example, Desvousges et al. (1993) attempted to test the theoretical validity which involves evaluations of how well contingent valuation estimates conform to hypotheses derived from a theoretical construct, using the migratory bird experiment. Three independent samples at two Malls in Atlanta, Georgia, were assigned three different scenarios that differ only in terms of number of bird death prevented from oil spills. That is, three versions of the survey (2000, 20,000 or 200,000) and their respective percentage (i.e. much less than 1%, less than 1% and above 2%) of the water fowl deaths prevented through protection of 250,000 ponds in the Central Flyway were presented to the respondents. The WTP values (of the open-ended results) for all the three treatments showed that there was no significant difference between the WTP values for these treatments indicating the presence of embedding effect. In addition to migratory water fowl scenarios, the authors also used oil spill scenarios (i.e. small spills vs. all spills) to see whether the theoretical validity of CVM for one commodity (i.e. migratory water fowl) is replicable for another commodity (i.e. oil spills). In the case of oil spill case too, the annual mean willingness to pay values for different levels of oil spill—derived from both open-ended as well as dichotomous choice techniques[10]—do not differ signifi-

---

[10] Later in this chapter, we discuss in detail various elicitation techniques including open-ended and dichotomous choice techniques.

cantly.[11] From these results, it has been concluded that the WTP values are infected with embedding effect, and therefore, the results of the CVM study are not theoretically valid (Desvousges et al., 1993). Boyle et al. (1994), using the results of migratory water fowl experiment of Desvousges et al. (1993), derive the following possible conclusions: (a) the marginal utility for preventing bird deaths (additional) is zero; (b) the marginal utility is greater than zero but is too small to detect; (c) the marginal utility is positive but not trivial, but CV is not able to measure the differences; (d) respondents' perception about the percentage of the birds prevented must have been subjective; and (e) the respondents WTP value must have been based on the number of ponds (over 250,000) to be covered which was same in all three treatment rather than number of bird death prevented (Boyle et al., 1994).

Having observed 'embedding' in CV studies that focused on measuring the non-use values, Mitchell and Carson (1989) conclude that embedding may be a problem only in the case of non-use values. However, Kahneman and Knestch (1992) argue that embedding is not restricted to non-use values alone but also occurs in the case of estimation of use values for public goods. It should be noted that embedding has been found to occur in CV studies on private goods. For example, Randall and Hoehn (1996) used two private goods—(i) a range of food items and (ii) rice [which is embedded in (i)] and demonstrated that the 'embedding effects observed with the contingent valuation of nonmarket goods are observed here with market demands for private goods' (p. 370). In another study, Bateman et al. (1997) used a similar kind of methodology on WTP for 'full meals' and WTP for 'part' of it and found the existence of embedding effect in their results. The conclusion of these two private good studies is that embedding is an ordinary economic phenomenon that not only occurs in public goods but also in private market goods. However, some of the studies found opposite results in which the embedding effect has been reported to be nil. For example, Choe et al. (1996) in their water quality study found that embedding effect did not occur for two different water quality levels that differed in scope.

It should be noted that different individual studies have come out with different kinds of results. But, what these studies have to say collectively? Smith and Osborne (1996) conducted a meta-analysis to assess the internal consistency of the CV results—especially the theoretical properties of the underlying preferences for the visibility improvements. This study utilises meta-analysis of five different CV studies that focused on changes in visibility at the US national parks. These five case studies were selected in such a way that they cover both use and non-use values and include both on-site surveys and surveys conducted in far away distance. Using Box_Cox and Feasible Generalised Least Square (FGLS), the mean WTP (MWTP) has been regressed against the independent

---

[11] In fact, the estimated means of the open-ended WTP values derived by using a different test for oil spill case show that the mean value for the small spill version is found to be greater than that of all spill version (Desvousges et al., 1993).

variables such as, eastern vs. western parks; decline or improvement in visibility; and the extent of the change (single park or the whole area). Other variables such as type of elicitation format used, whether interview took place within the park (or far away distance) and whether the households live in the park itself were also included in the analysis. Controlling for the sensitivity of the independent variables, the meta-analysis found that there exists a statistically significant, positive relationship between WTP and proportionate change in the visible range. Apart from the statistical significance between the WTP and the visible range, the analysis also investigated the proportionate change in the WTP value to the visible range and found existence of economic plausibility in the results. Overall, the meta-analysis reveals that the 'visibility changes do seem to be responsive to the NOAA panel's call for both statistical significance and economic plausibility in the scope test' (p. 299).

Though meta-analyses and individual studies provide theoretically consistent results, occurrence of scope effect (or embedding) in many other individual CV studies still causes disturbance. Having observed scope effect in CV studies, Diamond and Hausman (1994) provide two kinds of interpretation regarding this phenomenon. The first interpretation is that the CV studies valuing two commodities rather than one are unreliable. The other interpretation, as argued by Kahneman and Knestch (1992), is that the 'warm glow'[12] might have occurred in the CV results. It is, therefore, argued that presence of scope effect in the WTP/WTA accept values implies that the contingent valuation method may not be useful for estimating the non-use values (Kahneman and Knestch, 1992; Diamond and Hausman, 1994). This is because the occurrence of embedding effect violates the fundamental principle by which a rational consumer is directed to prefer more of a good or service to less and pay more for larger quantity compared to lesser quantity (Desvousges et al., 1993).

The proponents of the CV study do not agree with the argument by the opponents that CV method is not theoretically valid due to occurrence of embedding or scope effect. One of the reasons why the embedding shows up in some of the CV studies, according to Hanemann (1994), is due to simple economic theory. The marginal utility theory provides an answer to this phenomenon in the sense that individuals' utility at margin declines for the subsequent bundle of commodity they consume. However, empirical evidence not only shows that the difference in WTP for two different quantity levels is more or less equal but sometimes the decline in the value for the larger quantity is zero (NOAA, 1993), which is disturbing. But some of the authors argue that this is because of the way in which the commodity is being described in the CV studies (see Smith, 1992; Hanemann, 1994; Carson et al., 2001). For instance, scenario description that does not facilitate the respondents to make any real

---

[12] Warm glow occurs if the respondents get moral satisfaction from the act of giving (Diamond and Hausman, 1993).

difference between different levels of goods described, convenience sampling used in surveys (i.e. stopping the respondents at shopping malls and asking WTP questions in the case of Desvousges et al. (1993) study) and self-administered questionnaires that are prone to unreliable results, etc., are considered to affect the validity of the CV results (Hanemann, 1994).

Do the elicitation techniques used in the description of the CV scenario reduce the scope effect and improve the validity of the CV results? A comprehensive study by Hammitt and Graham (1999) used two different approaches to understand the influence of preference elicitation on the WTP values for reduction of health risk. The first approach involves meta-analysis of 25 CV studies on health-related aspects, such as road safety, medical treatment, hazardous waste management, etc. These studies were selected on the basis of three criteria, namely, (i): addressing the programme to protect and enhance the safety; (ii) eliciting WTP from lay respondents; and (iii) linking WTP value with the numerical change in the probability of death. These studies can be taken as to have used the conventional method of conveying different levels of health risk to the respondents. The results of the meta-analysis suggest that the sensitivity of WTP to probability variation in risk is limited. That is, the WTP values are found to be insensitive to the scope of the goods provided. In the second approach, the authors used a split sample test—consisting of 'internal' (within sample) and 'external' (between sample) tests of sensitivity to magnitude of risk. Two surveys have been used to estimate the WTP for risk reduction in two risk contexts, namely, transportation and food consumption. In Survey-1, the respondents were asked about only one risk, namely, dying in an automobile accident. One-half of the respondents were conveyed a baseline risk of dying in a road accident as 20/100,000 persons and the other half received a scenario with risk level of 25/100,000 person. Both the samples were asked about their WTP for purchasing a 'safety bag' that would reduce fatality risk to 15/100,000, and then to 10/100,000, from the baseline risk. In Survey-2 also, WTP for a safety bag was asked in addition to the food-risk scenario. But in Survey-2, an 'indifference-risk' approach is introduced to 'alleviate respondents' limited appreciation for the relative magnitude of numerical risk changes, a factor that may cause elicited WTP to be insensitive to magnitude' (p. 47). Here, innovations that would make the respondents understand the risk and to remove factors causing distortions between the preferences and the theory were used. The innovations include: (a) presentation of analogies in the stimulus to understand the respondents' understanding of risk; (b) distinguishing respondents with high and lower confidence levels; (c) information about the 'increments' in risk rather than 'levels' to combat the tendency to focus on percentage change; (d) use of one-and-a-half-bounded DC elicitation format in the food safety decision. In this survey, the respondents were asked to choose between two lunch stands—one is outdoor (cheap, more risky) and the other is indoor (more expensive, safer). In the case of high-risk scenario, risk levels of 1/13,700 and 1/137,000 were given to two subsamples,

while the low-risk scenario with only 1/100 million has been assigned to both the samples.

The results of Hammitt and Graham (1999) study provide certain useful insights. In the case of automobile survey, it was found that the indifference-risk approach (Survey-2) provided 'statistically significant difference' (between WTP values) than the conventional approach (Survey-1). In Survey-2, it was also found that the respondents with 'high confidence' in their answers provided WTP values that were proportional to the price difference than the respondents with less confidence. Overall, the authors conclude that the standard conventional CV methods do not provide theoretically consistent results. Nonetheless, the author's new approach that involved indifference-risk elicitation is found to provide theoretically consistent results. Therefore, the results of the meta-analysis and that of Survey-1 suggest that the conventional preference elicitation methods are generally weak in making the respondents to understand the differences in the level of commodity described. This implies that the scope effect or embedding could be addressed through proper 'study design'.

The findings of Hammitt and Graham's (1999) study are supported by their follow-up study reported in Corso et al. (2001). Corso et al.'s (2001) study has used visual aids to communicate two different levels of risk, namely, one in which the annual risk of dying in a motor-vehicle crash was 2.5/10,000 and the other in which it was 2.0/10,000. The information regarding the respondents willingness to pay for an automobile safety devise (i.e. a side-impact airbag) that would reduce the annual risk of dying in a motor crash to 1.5/10,000 was elicited through the 'mixed-mode phone–mail–phone survey' among four subsample groups of 1,104 final respondents. The findings suggest that the WTP values for independent subsamples presented with each of the three alternative visual aids (a linear scale, a logarithmic scale and an array of dots) were responsive to the magnitude of risk reduction. However, the WTP values elicited from the fourth subsample that received 'no visual aid' was not responsive to the magnitude. The conclusion of this study supports the notion that theoretically consistent CV results could be obtained through improved and effective communication devices.

One of the general conclusions regarding occurrence of scope effect in CV studies (that basically use conventional preference elicitation approach) on health risk is that WTP is not sensitive because, (i) respondents may not be sensitive to variation in risk magnitude due to their lack of understanding of probabilities and poor appreciation for numerical differences in magnitude (Kahneman and Tversky, 1973); (ii) respondents do not treat the probabilities as applicable to them and base their values on their prior beliefs and on the information contained in the scenario description. Finally, the respondents may not value the health risk in line with the expected utility theory (Hammitt and Graham, 1999).

To conclude, we have analysed different kinds of argument, supporting as well as rejecting the existence of embedding or scope effect in CV results. In the case

of studies reporting the existence of embedding, it is observed that the funda-
mental flaws in the design of a survey instrument, improper implementation of
the survey, improper sampling procedure, inability of respondents' understanding
of the survey questions, the properties attributed to standard value theory to
substantiate the claims for embedding, etc., are some of the factors that are found
to cause this problem (Hanemann, 1994; Smith, 1992; Harrison, 1992). Mitchell
and Carson (1989) suggest survey design features to minimise the potential for
part–whole biases. Describing the larger and smaller commodities, and then
asking respondents to focus their attention on the smaller commodity, using maps
and photographs to describe the scenario, debriefing, providing opportunity to
respondents to revise the bids, etc., are some of the measures suggested to
minimise the embedding in CV studies (NOAA, 1993; Mitchell and Carson,
1989).

2.3. Sequencing

Another source of error found to affect the value of WTP is the 'sequencing
effect' (Cummings et al., 1986), which is also called 'question order bias'
(Mitchell and Carson, 1989). Sequencing effect is said to occur if a WTP value
for a particular good differs depending on the order of the good in a sequence.
Like the embedding problem, the sequencing effect occurs in multigood
valuation studies (Diamond and Hausman, 1994).

There are empirical CV studies which have reported sequencing effect in their
results. For example, Samples and Hollyer (1990)[13] conducted an experiment
using two different goods—seals and whales to see the sequencing effect. Two
different scenarios—one with the WTP question for the seals first and the whales
next and another with the reverse order—were assigned to two different
respondent groups. The WTP value for the seals was found to be lower than
the whales' value for the scenario which described seals first and whales next,
and the value for the whales was found to be unaffected for the scenario that
described the reverse order. Kahneman and Knestch's (1992) study, discussed
under the embedding problem above, also found that the median WTP value for
'improved equipment, personnel' while valued alone, was found to be 25 times
greater than the median WTP value for the same good when valued in a sequence
of other 'environmental services'. Boyle et al. (1993) in a CV experiment
administered two water flow scenarios in the Grand Canyon among the white
water boaters. The total number of sample respondents were divided into two
equal groups. A scenario with seven levels of water flow in the Grand Canyon
was assigned to both the groups in such as way that the first group received the
scenario with a particular order while the second group received just the reverse
order. This study concludes that the sequencing effect was not identified for the
experienced boaters, whereas it was found to occur in some estimates for less

---

[13] Cited in Diamond and Hausman (1994).

experienced boaters. In other words, the 'familiarity' of the respondents with the 'good' under consideration plays a role in minimising the sequencing effect. In an attempt to estimate the influence of the people's perception about two risk levels on WTP, Hammitt and Graham (1999) found strong question order effect in the WTP values from two subsamples that received different order of the lower and higher risk levels. However, an earlier study by Ajzen et al. (1996) for a public good (construction of movie theatre in a college campus) and a private good (personal noise filter) in a laboratory experiment shows that WTP values for two sample groups that received two different scenarios with the changed sequence order of the private and public goods did not differ significantly.

According to Carson et al. (2001), the sequencing effect may be attributed to the influence of substitution and income effect (see also, Hoehn and Randall, 1989). More precisely, the respondents may substitute the first good for other goods in the sequence and will be left only with a little amount of money to spend on other goods after spending on the first good. Therefore, according to Carson et al. (2001), the sequencing problem is not due to the CV method as such but arises from the way in which the CV practitioner adds up the results, without taking into account the substitution and income effects. However, much larger sequencing effect which is not theoretically consistent is disturbing (see, McFadden, 1994), and therefore, there exists a need for minimising it. According to Mitchell and Carson (1989), the sequencing effect is another phenomenon that occurs due to improper administration of the survey. They suggest two procedures to minimise it: (i) informing the respondents about what is going to come before asking the WTP questions; and (ii) providing opportunity to the respondents to revise their bids after asking questions completely (also see Smith, 1992).

2.4. Information effect

The 'information' in a CV method plays a crucial role. The validity of the CV results depends mainly on the level and nature of information provided to the respondents through the scenarios. The nature of the information provided has been found to affect the results both positively as well as negatively (see Bergsrtom et al., 1990). The scenario in a CV study contains two major elements, namely, (a) the value-enhancing element (for instance, different levels of quantity of water supplied); and (b) the value neutral elements (such as photographs shown to describe, say, the visibility). Considerable amount of research has been carried out in the case of value-enhancing elements (e.g. Whitehead and Blomquist, 1990). The value enhancing-element in CV studies is of three types, namely, (i) the information about the good to be valued; (ii) the budget constraints and other peoples' CV values; and (iii) the information about the related environmental goods that are supposed to affect the WTP values for the good under consideration. For instance, information about related environmental goods that may be substitutes or complements has been found to influence stated willingness to pay. If substitutes are not presented, then the stated WTP becomes

an overstated one while the absence of a reminder of complementary goods leads to understate the stated WTP (Whitehead and Blomquist, 1990).

Does reminding of the budget constraint and substitute goods affect the WTP value? An empirical study by Whitehead and Blomquist (1990) looks into the impact of the information about the related environmental goods (substitutes and complements) on WTP values. The Clear Greek Wetland Systems (CGWS), the largest wetland tract in the Western Kentucky coal field, was chosen for the study. The aim of the study was to estimate the economic value of the CGWS—known for both its use and non-use values—which faces threat from a surface coal-mining project. Two related goods were chosen—a reclaimed lake as complementary good and an alternative wetland location as substitute good. Three versions of the scenario—each with a different set of information—were presented to three independent groups of sample respondents. A second dichotomous choice valuation question was presented with additional information about the related goods to all the sample households. Two kinds of valuation were possible with these two techniques: one is, three versions with different levels of information would provide information effect across different households, and second, the additional information to the same household would provide the incremental value for additional information included. The study found that information about the substitutes reduces the WTP while information about complements increases it. In Adamowicz et al.'s (1993) study that we have discussed in detail earlier, it was found that the information about the substitute goods has led to reduction (especially in the Hockey ticket treatment) in the disparity between the WTP and WTA values. Neill (1995) and Ajzen et al. (1996) also demonstrated that reminding of budget constraint and availability of substitutes influence the WTP values in laboratory experiments.

Some of the CV studies on the same issue have provided contradictory results as well. For example, Loomis et al. (1994) studied the impact of reminding budget constraint and substitutes on WTP values. This study used fire prevention and control program in 3 million acres of old growth forests in Northern Spotted Owl Critical Habitat Units in Oregon. The value to be elicited was non-use value. To test the information effect, two treatments were developed. In the first treatment, one half of the respondents received information regarding substitutes and the income constraints before the WTP questions and in the second treatment, the other half received remainder of substitutes and income constraints after the WTP question. The results showed that there was no difference between the WTP values under both the treatments implying that the information provided had no impact on the WTP values. The authors provide various interpretations: (i) the respondents might have already taken into account the budget constraints and substitutes while providing their WTP values; (ii) people might not have given sufficient consideration to the real dollar value especially when dealing with hypothetical scenarios; and (iii) respondents might not have thought the value of fire program in terms of dollar value.

The results regarding the information effect are mixed. But an important aspect to be noted is that the influence of the additional information on the WTP value depends mainly on the level of information possessed by the individuals. This implies that the CV studies should be capable of addressing the linkage between the levels of information possessed and the WTP value influenced by the additional information provided. In other words, if there exists asymmetric information across individuals, then the additional information provided in the CV studies should influence the WTP value as desired (Bergsrtom et al., 1990). Venkatachalam (2000) found that additional information provided about the drinking water quality to the respondents, who possessed different levels of information about the water quality, has significantly influenced the WTP values.

One crucial question that has not yet been properly addressed in the CV literature is that what is the optimum level of information, since the information provided in the CV scenarios do affect the WTP values. Given the fact that there is no well-defined answer for this question, one important aspect to note is that the information that should be provided in the scenario depends mainly on the intention of the CV practitioner, nature of the good to be valued, cost of obtaining information and so on.

2.5. Elicitation effects

The value of the good/service in the CV technique is elicited through an elicitation technique which is an important component of any CV method (Portney, 1994; Mitchell and Carson, 1989). The elicitation technique (or approach) used in CV studies is of different types. As of now, there are four major types of elicitation techniques available in the literature, namely, the bidding game, payment card (PC), open-ended (OE) and dichotomous choice (DC)[14] approach (Boyle et al., 1996). The dichotomous choice approach is further divided into two types as: single-bounded dichotomous choice or take-it-or-leave-it; and double-bounded dichotomous choice or take-it-or-leave-it with follow-up. An extended version of the latter approach which is called 'triple bounded dichotomous choice' that extends the double-bounded DC 'for a further question' has also been used in some of the CV studies (see Bateman et al., 1999). Let us discuss briefly each one of the elicitation techniques mentioned above.

The bidding game is the oldest elicitation technique among all the techniques (Mitchell and Carson, 1989). The bidding game approach goes as follows: the respondent in a CV study would be randomly assigned a particular bid from a range of predetermined bids. The bid assigned may be either a lower or higher level bid. The respondents would then be asked to say 'yes' or 'no' to that particular bid, and the process would continue until 'the highest positive response is recorded' (Randall et al., 1974, p. 135). Davis (1963) is the one who first used

---

[14] Also called 'referendum format'.

the bidding game approach for estimating the benefits from goose hunting. This approach has been later on used by many authors for estimating the value of public goods (e.g. Randall et al., 1974; Brookshire et al., 1982). Moreover, this is the approach which has been widely used in a relatively large number of CV studies conducted in developing countries (e.g. Whittington et al., 1990, 1992).

Using the bidding game approach for estimating the benefits from the abatement of aesthetic environmental damage associated with power plant and coal mine, Randall et al. (1974) conclude that the bidding game was successful in meeting the valuation objective. However, the authors suggest that considerable amount of care should be taken in designing the bidding game so as to obtain reliable results. The results of many of the developing country CV studies on water supply (Whittington et al., 1991, 1992; Briscoe et al., 1990) that utilised bidding game approach suggest that this approach works well in developing countries. One of the advantages of this approach is that it provides relatively better results since it gives a 'market-like' situation to the respondents in which they could research their preferences (Cummings et al., 1986). Another advantage of this approach is that the researcher could obtain maximum willingness to pay value (Cummings et al., 1986). However, Cummings et al. (1986) argue that the cost of implementing the bidding game is comparatively higher in the sense that it involves presence of interviewers during the interview, etc. Another problem with bidding game is that the starting points used in the bidding game might influence the final value of the stated WTP. Apart from these problems, Loomis (1990) argues that the bidding is impracticable in mail surveys.

The second oldest technique is payment card approach, introduced by Mitchell and Carson (1984). The payment card would contain a range of WTP values for the public good under question from which the individuals have to choose their maximum WTP value. The respondents are facilitated with another benchmark version of the payment card that contains the average WTP amount paid by households for other public goods. Even though the payment card approach has some advantages such as its ability to elicit the maximum willingness to pay value, the problem with the payment card is that there is a chance that the WTP values would be possibly affected by range bias and centering bias (Mitchell and Carson, 1989). Moreover, the payment card approach may have limited use especially in rural areas of developing countries where the people have very limited experience with using payment cards.

The open-ended elicitation technique involves asking what is the maximum amount that the individuals are willing to pay for a public good or policy. The open-ended approach is convenient to answer, does not require an interviewer and does not result in any starting point bias (Walsh et al., 1984). For those studies which aim at deriving a value that would provide a conservative estimate, the open-ended approach would be efficient in the sense that this approach would provide a lower level conservative value than the bidding game approach (Walsh et al., 1984). However, this approach is prone to criticisms. Desvousges et al. (1993) point out that the open-ended approach tends to create large number of

nonresponses or protest bids since respondents either find it difficult to answer or do not have incentive to provide true answer (Carson et al., 1996). Hanemann (1994) argues that the open-ended questions may attract strategic bias and people may tell the cost rather than true value.

Due to problems with the techniques mentioned above, Bishop and Heberlein (1979) introduced take-it-or-leave-it approach (or single-bounded dichotomous choice approach). It involves assigning a single bid from a range of predetermined bids that potentially reflect the maximum willingness to pay amounts of the respondents for a particular good. The respondents are asked to state only 'yes' or 'no' to that bid on all or nothing basis (Mitchell and Carson, 1989). The major advantage of the single-bounded dichotomous choice approach is that it facilitates the respondents to complete the valuation process. Moreover, the strategic bias in the WTP values may be minimised since this approach is an incentive compatible one (Carson et al., 1996; Hanemann, 1994). Despite its advantages, the single-bounded dichotomous choice approach has some disadvantages as well. One of the disadvantages is that one can derive only the maximum willingness to pay/minimum willingness to accept value from this approach but not the actual willingness to pay amount (Boyle et al., 1996). This approach also attracts starting point bias (Ready et al., 1996). On the property of incentive compatibility, this approach may not work in two circumstances: when the public good is provided through voluntary contribution and when a new private or public good is provided (Carson et al., 1996). Moreover, this approach is less applicable in areas where the households have already decided about their willingness to pay value (Venkatachalam, 2000). Another problem is that it requires a large number of observations for identifying the distribution of values (Alberini, 1995; Cameron and Quiggin, 1994).

A modified version of the take-it-or-leave-it approach has been introduced by Hanemann (1984, 1985) and Carson (1985) which is called, 'double-bounded dichotomous choice approach' (or 'take-it-or-leave-it-with follow up'). This involves assigning one more bid to the initial bid (in the single-bounded approach), whose direction depends on the 'yes' or 'no' answer to the initial bid. This approach has been first applied by Carson and Steinberg (1990) and Hanemann et al. (1991). The major advantage of this approach is that one could identify the location of the maximum willingness to pay value from the data derived from this approach. This approach is an incentive compatible one. The double-bounded dichotomous choice approach is statistically more efficient than the single-bounded dichotomous choice approach (Kanninen,1993; Hanemann, 1991). Conducting an econometric analysis of the simulated data, Alberini (1995) concludes 'to obtain more powerful goodness-of-fit tests so that a poor specification of the model can be detected and corrected in the early stages of the data analysis, the CV researchers are advised to plan discrete choice surveys with a follow-up bid …' (p. 95). A problem with this form of dichotomous choice approach is that it requires a larger sample size, sophisticated econometric techniques, etc. which results in increased cost of the survey. Moreover, the

results derived from the double-bounded dichotomous approach are vulnerable for starting point bias and 'yea-saying' problem (Ready et al., 1996).

2.5.1. Comparison of elicitation techniques

In the CV literature, it is generally reported that using different elicitation techniques provides different kinds of WTP values, and therefore, one is confused whether the WTP values are true values or arising from the type of elicitation technique used. This being the crucial issue, different studies compare different elicitation techniques mainly to test the 'convergent validity' of the CV results. These studies derive different kinds of conclusions as well. In the oil spill case discussed above, Desvousges et al. (1993) used both open-ended and dichotomous choice techniques for two versions of oil spill—small spills and all spills and found that '. . .the dichotomous format yields WTP estimates that are statistically significantly larger than the comparable estimate from the open-ended format' (p. 106). Kealy and Turner's (1993) study on measuring the WTP values for a private good (a candy bar) and a public good (preventing acid rain deposition) reveals that there is no statistical difference between results derived from open-ended and DC techniques for the private good but a significant difference is found in the case of the public good. Brown et al.'s (1996) study found that the WTP value from the DC technique is greater than that of the OE. A study on WTP for 'food safety improvements' that used dichotomous choice and open-ended techniques to independent samples by Ready et al. (1996) reveals that the dichotomous choice technique has generated a much larger estimate than the open-ended technique. Using Monte Carlo experiment to compare WTP values from referendum format and payment card approach, Jordan and Elna-gheep (1994) conclude that the parameter estimates from the payment card experiment are more efficient than the referendum experiment. Holms and Kramer (1995), in attempt to test the convergent validity of the CV method by looking into the values derived from two elicitation techniques, namely, payment card and dichotomous choice, conclude that 'yea-saying'[15] and starting point bias might have been the reasons for higher WTP value in the DC approach. Schulze et al. (1996) review various studies that look into the difference between DC and OE and the review suggests that DC/OE ratio ranges between 1 and 72.9. A similar review by Brown et al. (1996) suggests that the DC/OE ratio falls between 1 and 5, for both private and public goods.

In recent years, improved versions of elicitation techniques have been used in different contexts to test the convergent validity. In a comprehensive experiment, Welsh and Poe (1998) compared results obtained from multiple-bounded discrete choice (MBDC) model 'which allows respondents to vote on a wide range of referendum thresholds' with the WTP values elicited through payment card, dichotomous choice and the open-ended formats assigned to separate subsamples

---

[15] Yea-saying refers to 'the tendency of some respondents to agree with the interviewer's request regardless of their true views' (Mitchell and Carson, 1989, pp. 240–241).

each. In the MBDC approach, the respondents in a subsample were provided with 'polychotomous' choice option that facilitates the respondents to gain level of voting certainty for a particular dollar threshold. For example, the respondents were asked to state any one of the five answers, namely, 'definitely no', 'probably no', not sure', 'probably yes' and 'definitely yes' for each dollar threshold. The results of this study suggest that inferences from OE, PC and DC fall within the range of MBDC estimates. A comparison between MBDC and the single-bounded DC values suggest that, ceteris paribus, when a respondent is uncertain about his payment obligation for a single bid, she will have the tendency to say 'yes' to the bid. However, the WTP measures elicited through OE and PC are found to be consistent with the higher level of certainty. Since it has been found that the CV values resulting from higher level of certainty are valid, this study suggests that 'definitely yes' value of the MBDC model is more appropriate for decision-making.

Cameron et al. (2002) made a comparison of choices elicited through seven alternative preference elicitation methods, consisting of one revealed preference and six stated preference elicitation formats. Dichotomous choice with single bid, dichotomous choice with varying bids, open-ended, payment card, multiple-bounded and the conjoint analysis elicitation formats have been used among the split samples. One telephone survey and six mail surveys have been administered to the sample households who were asked to consider (either hypothetically or in real terms) additional charges on their utility bills due tree planting and/or generating energy from renewable sources by Niagara Mohawk Power Company. The results of the study suggest that the underlying preference functions across different elicitation formats differ when each elicitation format is treated alone. However, when different types of WTP values are pooled, the underlying systematic preference functions were found to be identical for four out of all seven elicitation formats used. It was found that the open-ended and payment card methods were found to be least consistent with other methods.

Many studies have provided results which suggest that, in general, the WTP value elicited using DC is greater than that of OE method (Carson et al.1996). This phenomenon is attributed to various reasons: (a) occurrence of strategic bias (especially, understatement) in OE, whereas this is not the case in DC which is considered to be incentive compatible; (b) 'yea-saying' in the case of DC format; and (c) tendency of the respondents to provide a lower WTP value when faced with a more difficult open-ended WTP questions (see, Brown et al., 1996). In recent years, a shift from using OE to DC is taking place in the CV literature but this shift demands for a sophisticated statistical analysis of the CV data. Modeling the DC data within the framework of random utility model (RUM) is a rigorous exercise and different kinds of DC formats require different kind of statistical analysis. Hanemann and Kanninen (1999) provide excellent description about various models, which could be used for analysing the DC data to arrive at relevant measures.

It should be noted from the above discussion that different elicitation techniques have got different kinds of advantages and disadvantages. This being the case, the question one has to address in a CV survey is which one of these techniques should be used to elicit the value of public goods. Mitchell and Carson (1989) report that the open-ended method works smoothly in situations where the respondents are familiar with paying for the goods under question while many others conclude that DC approach is more incentive compatible than other elicitation techniques especially in the case of non-use values (NOAA, 1993; Hanemann, 1991; Hanemann and Kanninen, 1999). However, one of the important questions still needs to be answered is: if the value differs among different elicitation formats, then what to do? Hanemann and Kanninen (1999) provide the answer by arguing that the cognitive demands of the individuals are not identical, and therefore, one should not expect that the values across different elicitation techniques should converge. In nutshell, selection of an elicitation technique in a CV survey depends on different factors such as the nature of the good investigated, cost of the survey, nature of the respondents targeted, nature of the statistical technique used, etc. As Hanemann (1994) argues, if people carry utility function in their brains, question format does not matter in a CV study.

2.6. Hypothetical bias

The nature of the market created in a contingent valuation survey is mainly hypothetical, and therefore, it may attract a bias called 'hypothetical bias' (Neill et al., 1994). This bias is defined as the potential divergence between the real and hypothetical payments (Cummings et al., 1986). Many CV studies have reported that the hypothetical WTP values are found to be greater than the real WTP values (Brown et al., 1996; Neill et al., 1994; Kealy et al., 1990; Bishop and Heberlein, 1979). For instance, Duffield and Paterson (1991) in an experiment estimate the WTP for maintaining the river flow that would facilitate protection of two rare fish species. Two independent samples were used to estimate the non-use value of the fish species in this case. Respondents in one sample group were asked to state their hypothetical WTP for the Montana Nature Conservancy—a body that would maintain the stream flow in the river and respondents in the other sample group were asked to actually contribute to the same organisation. The results of this study show that the amount of hypothetical WTP exceeds the actual WTP. In another study by Seip and Strand (1992), hypothetical WTP value was elicited from a sample group for membership fee for a Norwegian environmental organisation. The same sample group then was asked to contribute 'actual payment' towards the membership fee. In this case, it is reported that the hypothetical WTP value was greater than the actual contribution. Neill et al. (1994) conducted an experiment involving two private goods, namely, (i) a water colour painting by a Navajo Indian artist and (ii) a framed print 16th century map of the world, to see the relationship between the WTP amounts elicited by the hypothetical contingent valuation questions and the actual payments if the

respondents are provided with the good. The authors conclude that the hypothetical WTP is significantly and consistently higher than the actual WTP. Foster et al.'s (1997) study compares the actual donations to environmental preservation[16] and the hypothetical WTP values derived from six UK CV studies for comparable environmental amenities.[17] The important finding of the study is that there exists a divergence between the actual and hypothetical WTP values, especially the hypothetical value being greater than the actual value. The results of an earlier study by Brookshire and Coursey (1987) indicate that the WTP value of a hypothetical CV market for expansion/reduction in the tree cover in a particular locality is greater than that of the actual laboratory experiment for the same issue. Moreover, the repeated trials in the laboratory experiments show that the difference between the WTP and WTA values in the hypothetical market situation converged to a greater extent (from 75 times to 5 times). This implies that the repeated trials make the respondents familiar with many aspects of the CV market, reducing the hypothetical bias to a considerable extent. In their own split sample experiment, Brown et al. (1996) used dichotomous choice and the open-ended techniques for both hypothetical and actual payment scenarios. The WTP value was elicited for the cost of meeting the food and supplies of voluntary labourers who would remove the abandoned unpaved roads in the Grand Canyon region. The results of the study revealed that in the actual payment scenario (where the respondents were asked to write a cheque to the 'Grand Canyon National Park Service'), the mean WTP for DC was $7.22 and that of OE was $4.62, the difference being 1.6 times. In the hypothetical condition, the mean WTP values for DC was $47 and that of OE was $19—that is, the DC/OE ratio being 2.5 times.

However, there are studies which have found the stated WTP value to be an understatement of the actual willingness to pay value. For example, Bishop and Heberlein's (1979) hypothetical and simulated experiments for goose-hunting permits reveal that in the simulated experiment, the actual willingness to accept value for a permit (i.e. the respondent's selling price of her permit) was $63, whereas in the hypothetical market, the WTP value was $21 and the WTA value was $101. Treating the actual WTA value in the simulated market as the true value, Bishop and Heberlein (1979) conclude that, in the hypothetical market, the WTP value is understatement of true WTP and the WTA value is overstatement of it.

Except few, many studies report hypothetical WTP value to be higher than the actual WTP value. It should be noted that many of these studies are found to be laboratory experiments that involve mainly private goods. There are some

---

[16] It involves donation to six large-scale fund rising appeals in UK, namely, (a) land purchase of maritime heath habitat; (b) land purchase of low-lying wet grassland; (c) campaign to protect coastal habitats for shorebirds; (d) campaign to protect reed-bed habitat for bittern; and (e) preservation of lowland heath habitat.

[17] Especially for three goods, namely, heathland, wetland and ancient woodland.

problems with the laboratory experiments. The first problem is that since students involved in these studies are not true representatives of general population, these experiments would become irrelevant and misleading.[18] The second problem is that in many of these experiments, mainly the private goods are used, and therefore, the results may be less relevant for addressing the issues of public goods. Apart from these criticisms, one more aspect to be noted is that in laboratory experiments the subjects in most of the cases are provided some money prior to the survey so that they could not experience any 'budget constraint'. But in actual markets, the individuals' WTP is constrained by the income. Moreover, results of many CV studies suggest that the 'familiarity issue' plays a dominant role in minimising the hypothetical bias in CV studies. More precisely, the more a respondent is familiar with the good, the less will be the level of hypothetical bias in a CV method (Mitchell and Carson, 1989). This implies that the WTP values elicited for those public goods, which are traded in the markets or which the individuals are familiar with, would be free from hypothetical bias (Whittington et al., 1991).

2.7. Strategic bias

The strategic bias is another problem in CV studies. There are two forms of strategic behaviour, namely, free riding and overpledging (Mitchell and Carson, 1989). Free riding would occur if an individual understates her true WTP for a public good on the expectation that others would pay enough for that good, and therefore, she need not have to pay. On the other hand, overpledging occurs when an individual assumes that her stated WTP value would influence the provision of good under question, provided that the stated WTP would not form any basis for the future pricing policy.

The debate on strategic bias was initiated by Samuelson (1954) when he criticised the survey methods on the ground that individuals would always tend to 'free ride' on survey questions. Following this, many laboratory experiments have been conducted and the experiments provide mixed results about occurrence of strategic bias in CV results (see Mitchell and Carson, 1989). Some of the empirical studies have also reported strategic bias in their results. For example, Whittington et al. (1992) designed a study to test the impact of time given to the respondents on WTP value for improved services of water supply. Two independent sample households were selected from three villages of Nigeria. One set of sample households was given an opportunity (1-day time period) to 'think' about their WTP value, whereas the other set of sample households was not given this opportunity. The results suggest that the WTP values for improved water services—elicited from those households who were provided time to think about their stated WTP—are found to be less than that of the households who

---

[18] However, results of the meta-analysis by Horowitz and McConnell (2002) do not support this claim.

were not provided the time to think. This implies that the households who were given time to think about the WTP values might have behaved strategically (by understating their true WTP) on the assumption that their stated WTP value would form the basis for the future water tariff policy.[19]

It should be noted that there are only very few CV studies that exclusively deal with addressing the issue of strategic bias. Many of the CV studies take a stand that the strategic bias is not a major problem in CV experiments (Griffin et al., 1995; Schulze et al., 1981). According to Mitchell and Carson (1989), the following reasons make the strategic behaviour very weak for most of the CV respondents: (a) the amount of information required for strategic behaviour are great; (b) CV surveys convey to the respondents that a larger number of people are interviewed, and therefore, respondents get the impression that their stated WTP would not influence the overall outcome; (c) the payment vehicles used in CV studies remind the respondents about the budget constraint so that the respondents could not overstate their true WTP; and (d) the understatement of true willingness to pay might be discouraged given the respondents' impression that the good under investigation may not be provided. Apart from these aspects, it has been found that using incentive compatible elicitation techniques (such as dichotomous choice technique) would minimise the impact of strategic bias (see Carson et al., 2001). Having reviewed different kinds of experiments on strategic bias, Mitchell and Carson (1989) suggest that the CV questionnaires should be designed such that it would not give any 'hint' to the respondents that makes them behave strategically.

2.7.1. Reliability of CV results

The reliability of the CV results is assessed in terms of: (a) testing for the convergent validity; and (b) comparing the WTP values through test–retest method (Hanley et al., 1997). The convergent validity of the CV method is assessed by way of comparing the CV results with the values obtained through another revealed preference method (Carson et al., 1996). Brookshire et al. (1982) compared the results of the CV method with that of the hedonic pricing method for air quality improvement. Their assumption is that values derived from the hedonic pricing method should be greater than that of CV results.[20] The results of their analysis show that the CV results are comparatively lower than the results derived from the hedonic pricing method, as expected. John et al. (1992) compared the results of CV study with the results of the expenditure function approach for a mosquito control program in Jefferson County, TX. The benefits

---

[19] However, the authors claim that the respondents provided genuine reasons for their comparatively lower level of WTP values, in an informal discussion after the main survey was over. This, according to the authors, implies that it is not the strategic bias but other factors, such as socio, economic variables, etc., affected the WTP values.

[20] This is because if the households' WTP value is greater than the hedonic pricing value, this suggests that the households should have already moved to unpolluted area rather than staying in their present locality.

of mosquito control program derived from expenditure function approach were found to be greater than that of the CVM. Carson et al. (1996) conducted a meta-analysis of 83 studies that made 616 comparisons of CV results with results of revealed preference methods. These studies have compared CV value with the value of any one of revealed preference methods—single-site travel cost method (TC1) (295 comparisons), multisite travel cost method (TC2) (183 comparisons), hedonic pricing method (HPM) (62 comparisons), averting expenditure method (28 comparisons) and actual market price or simulated market (48 comparisons). The comparison between the CV value and the value of a revealed preference method has been made for essentially the same quasi-public good. The goods covered in the studies are: outdoor recreation, changes in health risks and changes in environmental amenities such as air pollution, noise pollution, water pollution or parks. The meta-analysis covers only those studies which used WTP method, but not WTA. Another feature of this meta-analysis is that it has covered both published and unpublished literature such as journal articles, theses, dissertations, conference papers and government reports right from 1966 to 1994. The results of the meta-analysis show that CV estimates are found to be 20% lower than TC1, 30% lower than TC2, little less than 40% to HPM, 20% to averting method and indistinguishable to actual market value.

Having reviewed the studies comparing the CV results with that of other alternatives, Smith (1993) concludes '…some forms of CVM can provide theoretically consistent and plausible measures of individuals' values for some types of environmental resources' (p. 14). However, one of the problems with comparing the CV results with the results of revealed preference methods is that the results of the revealed preference methods themselves, in many cases, are found to be a poor indicator of true value of the environmental resources. Therefore, comparing the CV results with another result that is already poor may not be adequately reflecting the issue. Moreover, the issues addressed, the benefit measured, the nature of the market used, etc., in the CV method are different from other revealed preference methods.

The second type of reliability test, namely, the 'test–retest method' involves conducting CV for a particular issue among a particular sample households and then repeating the same study either among the same sample in two time periods or among different sample households in the same population (see Hanley et al., 1997). Using a private good (candy bars) and a public good (de-acidification of lakes), Kealy et al. (1990) tested the reliability of WTP values obtained from open-ended and dichotomous choice techniques using test–retest method. The time gap between the first test and the second test was 2 weeks. The difference between the WTP values of both the techniques has been found to be statistically insignificant in both the time periods, which suggests that the results obtained are reliable. However, too short of a time gap between the two tests is supposed to have resulted in a recall effect, i.e., respondents recalling their answer in the previous test (Teisl et al., 1995). This leads to another question, namely, whether the time gap between two studies could explain the conver-

gence between the WTP values. Loomis (1990) used a test–retest procedure for open-ended and dichotomous choice techniques for different levels of water quality in Mono Lake. The time gap between the two tests in this case was 9 months. The results of this study show that there is no statistically significant difference between the values obtained through open-ended and dichotomous choice techniques in two time periods. Teisl et al. (1995) conducted a test–retest procedure using open-ended and dichotomous choice techniques for moose-hunting permits among sample respondents with and without the moose-hunting experience with a time gap of 5 months. The authors conclude that the WTP estimates are found to be reliable for both experienced and nonexperienced moose hunters. In general, a majority of the studies that conducted the reliability tests conclude that the contingent valuation method is capable of generating reliable WTP results.

One of the important issues still to be resolved in the case of test–retest method is that what should be a 'reasonable' time gap between the first test and the second test so that any recall effect or the impact of larger changes in the socioeconomic variables of the individuals/households could be addressed properly in the analysis. Another problem is that one particular elicitation technique is assumed to be the 'perfect' measure of value so that the other one could be compared with the perfect one for assessing the reliability of the results (see Smith, 1992). However, the empirical evidence shows that different elicitation techniques have different kinds of problems and no single elicitation technique has been proved to be perfect in eliciting the true economic value.

2.7.2. Benefit transfer

The reliability issue is closely inter-linked with another issue, namely, benefit transfer (BT). More precisely, the issue being addressed under benefit transfer approach is that whether preferences estimated from a study site could be transferred to a policy site to predict the behaviour of the respondents (see, Bateman et al., 2002). Some of the empirical studies have looked into this issue. The major objective of Downing and Ozuna's (1996) study is to test whether the statistically similar contingent valuation benefit transfer functions yield statistically similar welfare measures. The test is based on the data collected from saltwater anglers surveyed in eight Texas Gulf Coast bay regions in three different time periods. The findings of the study suggest that many of the benefit functions between within and across bay for two time periods are transferable. However, the transferred benefit functions could not yield statistically similar welfare measures and the study, therefore, concludes that the benefit function transfer approach that transfers benefits from a study site to a policy site could be misleading. According to the authors, one of the reasons for the results might be the 'nonlinearity model' used for deriving the results. A similar study by Kirchhoff et al. (1997) to test the performance of both benefit transfer and benefit function transfer regarding preference for nonmarket amenities found that the benefit function transfer is found to be more robust than the benefit transfer

method. The authors conclude that the errors occurring in the application of benefit transfer method may be quite large.

In a developing country context, Griffin et al.'s (1995) study investigated two issues of CV method, namely, (i) benefit revelation issue (i.e. Is the people's behaviour same in both ex-ante and ex-post situations?) and (ii) benefit transfer issue (i.e. Is it possible to transfer valuation from one population to estimate how a second population would value the same resource?). This is a test–retest-type study conducted during 1988 (ex-ante) and 1991 (ex-post), in northern areas of the Indian state of Kerala. The 1988 study was carried out among the households in three different drinking water environments, namely, abundant water area, scarce water area and an area with salt water intrusion. In each category, two types of households—site A community households and site B community households—were chosen. In the case of site-A community households, piped water services were already available to these households. These households were further classified into two categories: (i) those who had already decided to connect to individual house connections at the existing connection costs and tariffs, and (ii) those who decided not to connect. Households in the site B communities had no piped system but were expecting to connect to such a system. The results of 1998 survey were compared with that of the 1991 survey. In the case of scarce water site, the 1998 survey's prediction that 14.2% of the total households would connect to house connections was consistent with the results obtained from the actual behaviour of the households. The 1991 resurvey has found that 91% of the households out of all the households who said they would connect to individual connections had actually connected in 1991. In the case of remaining households in this category which had not connected, more than 75% of the households reported that inability to pay the connection cost to be the major reason. In the case of households who actually connected but who were not willing to connect during 1988 survey said 'changed economic circumstances' as the major reason for their decision. The results of the scarce water A site had been used to predict the people's behaviour in the B site and it was found that site B provided 'completely inaccurate predictions'. Therefore, it has been concluded that benefit transfer approach was not accurate.

Some of the earlier studies have also concluded that the benefit transfer method is unreliable (e.g. Kirchhoff et al., 1997). Benefit transfer method is found to work well in some contexts and not in other (Bateman et al., 2002). This may be attributed to the fact that the respondents in one site may not be a perfect substitute for the respondents in another site due to differences in socioeconomic characteristics of the respondents. Since the benefit transfer approach is still in its infancy level, more research is needed on this topic.

So far, we have been discussing various major issues that affect the CV results and how these issues are addressed in different CV studies. It should, however, be noted that the list of issues we have discussed is not exhaustive. Mitchell and Carson (1989) have listed various kinds of biases/errors that are supposed to affect the CV results. Since the CV method attracts these kinds of biases and

errors, the critics of the CV method argue that this method is not capable of estimating the economic values, especially the non-use values (see Hausman, 1993). On the other hand, the supporters of the CV method argue that the problems associated with the CV method are mainly due to poor administration and implementation of the CV survey, and therefore, the CV method is capable of eliciting the true economic values of the environmental resources if properly conducted (Mitchell and Carson, 1989; NOAA, 1993; Portney, 1994; Hanemann, 1994). The supporters of the CV method, in this direction, have developed many different guidelines and procedures for conducting the CV method that could elicit valid and reliable results. The following section discusses these guidelines in detail.

3. Guidelines for conducting CV studies

Cummings et al. (1986), having reviewed various issues involved in the CV method, have formulated some guidelines described as the 'reference operating conditions'. The five reference operating conditions are: (i) use familiar environmental goods; (ii) respondents should have some valuation experience of the environmental change in question; (iii) the scenario should not have a high degree of uncertainty; (iv) WTA scenarios should not be used; and (v) use values are likely to be more accurate than non-use values.

Following Cummings et al. (1986), Bateman and Turner, (1993) have provided a comprehensive list of guidelines. They include: (i) only apply CVM to goods with which the respondents have some familiarity; (ii) scenario should be realistic, plausible, clearly understood and not have a high degree of uncertainty; (iii) WTA scenarios should be avoided; (iv) the payment vehicle should be realistic and appropriate; (v) estimates of use values are likely to be more accurate than the non-use values; (vi) use both open-ended and dichotomous choice formats to provide lower and upper valuation boundary estimates; (vii) the survey should question intentions to behave rather than attitudes towards behaviour; (viii) the scenario should make provision of the good dependent upon behaviour; (ix) adequate (rather than excessive) unbiased information should be provided, with the impact of that information assessed via a control group who receive no information; (x) specific questions should be included to minimise part–whole (mental account) problems; (xi) sample size must be statistically significant; (xii) avoid starting points; (xiii) avoid any direct or implied-value cues either via information, questionnaire or interviewer; (xiv) choose carefully between face-to-face and remote (mail, etc.) approaches and ensure that the correct population is being sampled; (xv) CVM will work best where respondents have some experience of valuing the good in question; (xvi) in dichotomous choice format, the upper bid should be selected such that almost 100% rejection is achieved, while the lower bid level should achieve almost 100% acceptance; (xvii) analysts should consider carefully the removal of outliers and the use of

trimmed means; (xviii) telling respondents that payments by others will be compulsory and may reduce nonresponse but may increase free-riding and strategic bidding; (xix) great care is required in the aggregation process; (xx) theoretical validity testing (bid function estimation) should be carried out. Further validity tests should be included where feasible; (xxi) where possible, assess reliability by retesting at a later date; and (xxii) report in full all results, including all sample statistics, details of information given and a reprint of the full questionnaire.

The NOAA panel constituted to evaluate the use of contingent valuation method in estimating the non-use values concluded that the CV could be used to derive useful information about the non-use values provided that the panel's guidelines are followed while conducting CV study. Some of the important guidelines are: (a) using probability sampling for damage assessment studies; (b) minimising nonresponse; (c) using personal interview rather than mail surveys and telephone interviews; (d) pretesting for interviewer effect; (e) reporting about the definition of population sampled, sampling frame used, sample nonresponse rate, etc; (f) pretesting the CV questionnaire; (g) using conservative values while the responses are ambiguous; (h) using the WTP format rather than WTA format; (i) using the referendum format rather than open-ended; (j) describing the programme or policy accurately to the respondents; (k) pretesting the photographs used in describing the scenarios; (l) reminding the respondents about the availability of the undamaged substitutes; (m) making sure adequate time lapse between the survey and the environmental insult so that the respondents find complete restoration of the damage plausible; (o) eliciting the reasons for yes/no answers for the valuation questions; and (p) reminding of the respondents of alternative expenditure possibilities.

It should be noted that these guidelines are general in nature and applicability of these guidelines, particularly in developing countries, depends mainly on the socioeconomic–institutional aspects prevailing in the study area. Therefore, the CV practitioners will have to be selective in using these guidelines.

4. Conclusion

It should be noted that nowadays the CV method has become an integral part of environmental assessment of developmental and basic infrastructural projects. However, it has been noticed that a considerable amount of CV studies is being indiscriminately conducted, without assessing the validity and reliability of the results. Since different kinds of biases and errors would potentially affect the CV results which may be less relevant for policy-making purpose, it is strongly recommended that all the CV studies should be designed such that the 'internal consistency tests' (Smith and Osborne, 1996) could be carried in a later stage to assess the validity and the reliability of the results. Moreover, review of a considerable number of CV studies suggests that some of the biases discussed

above occur due to improper administration and implementation of the CV method (Carson et al., 2001), and therefore, pretesting the interview schedule and pilot studies should be made a precondition in any CV study. This would reduce certain biases in the preliminary stage of the survey itself. It is also recommended that if the guidelines provided by various authors are followed carefully, many of the potential biases could be either controlled or minimised. However, as we have already seen, these guidelines are only general guidelines in nature and do not provide any concrete design or framework for conducting an ideal CV survey. The experience with various CV experiments suggests that the specific guidelines should emerge from the field and not from the texts and in developing countries especially, where the socio, economic and institutional setup pose certain difficulties to conduct a good CV study, application of these guidelines needs some amount of scrutiny. For example, the NOAA (1993) panel suggests that the WTP measure compared to the WTA measure is a suitable measure of value, and therefore, the WTP rather than the WTA measure should be used in CV studies. However, in many of the developing countries, the environmental impact assessment studies look at the negative impact of the developmental and environmental projects on the poorer section of the society who are indirectly paying in terms of damage cost. In this case, from the point of view of the losers, the WTA compensation rather than WTP is considered as an appropriate measure.

Another important aspect to be noted is that the scenario in the interview schedule should accommodate as much information as possible from the respondents during pretesting. This would make the scenarios more meaningful and understandable to the respondents during the main survey. This is because in almost all the CV studies, the information provided in the scenario consists mainly of the information derived from the 'intention' of the researchers. All that the respondents have to do is to either accept the scenario or reject it. This would ultimately result in a large number of protest bids. This kind of approach, however, may not be relevant because the cost–benefit analysis or the environmental impact assessment exercises always look for information that would be useful for important policy-making purpose such as fixing-up of tariff for basic infrastructural facilities, setting compensation for the losers of an environmental project, etc. Therefore, there exists a need for enhancing the scenarios with the information from the respondents, who are largely the gainers or the losers of an outcome of a particular decision.

To conclude from Carson et al.'s (2001) words, 'Even if all of the survey related issues to valuing a public good can be overcome, CV is not without its limitations' (p. 197). It may be observed that the major implication of our review is that even though CV had certain limitations, this method is a promising method and it could be used to derive useful information. But this does not mean that this method could be used indiscriminately. It is strongly recommend, in line with the recommendations of the CV literature, that the CV study should be properly and extremely carefully conducted. Otherwise, the results of the CV method would lead to misleading conclusions both in the academic and policy-making arenas.

Acknowledgements

References

Adamowicz WL, Bhardwaj V, Macnab B. Experiments on the difference between willingness to pay and willingness to accept. Land Economics 1993;69:416–27.

Ajzen I, Brown TC, Rosenthal LH. Information bias in contingent valuation: effects of personal relevance, quality of information, and motivational orientation. Journal of Environmental Economics and Management 1996;30:43–57.

Alberini A. Testing willingness-to-pay models of discrete choice contingent valuation survey data. Land Economics 1995;71:83–95.

Bateman IJ, Turner RK. Valuation of environment, methods and techniques: the contingent valuation method. In: Kerry Turner R, editor. Sustainable environmental economics and management: principles and practice. London: Belhaven Press; 1993. p. 120–91.

Bateman IJ, Munro A, Rhodes B, Starmer C, Sugden R. Does part–whole bias exist? An experimental investigation. Economic Journal 1997;107:322–32.

Bateman IJ, Langford IH, Rasbash J. Willingness-to-pay question format effects in contingent valuation studies. In: Bateman IJ, Willis KG, editors. Valuing environmental preferences. Oxford: Oxford Univ. Press; 1999. p. 511–39.

Bateman IJ, Carson RT, Day B, Hanemann MW, Hanley N, Hett T, et al. Economic valuation with stated preference techniques: a manual. Cheltenham: Edward Elgar; 2002.

Bergsrtom JC, Stoll JR, Randall A. The impact of information on environmental commodity valuation decisions. American Journal of Agricultural Economics 1990;72:614–21.

Bishop RC, Heberlein TA. Measuring values of extra-market goods: are indirect measures biased? American Journal of Agricultural Economics 1979;61:926–30.

Boyle KJ, Welsh MP, Bishop RC. The role of question order and respondent experience in contingent-valuation studies. Journal of Environmental Economics and Management 1993;95:S80–90.

Boyle KJ, Desvousges WH, Johnson FR, Dunford RW, Hudson SP. An investigation of part–whole biases in contingent-valuation studies. Journal of Environmental Economics and Management 1994;27:64–83.

Boyle KJ, Johnson FR, McCollum DW, Desvousges WH, Dunford R, Hudson S. Valuing public goods: discrete versus continuous contingent-valuation responses. Land Economics 1996;72: 381–96.

Briscoe J, de Castro PF, Griffin C, North J, Olsen O. Toward equitable and sustainable rural water supplies: a contingent valuation study in Brazil. The World Bank Economic Review 1990;4:115–34.

Brookshire DS, Coursey DL. Measuring the value of a public good: an empirical comparison of elicitation procedures. American Economic Review 1987;77:554–66.

Brookshire DS, Thayer MA, Schulze WP, d'Arge RC. Valuing public goods: a comparison of survey and hedonic approach. American Economic Review 1982;72:165–76.

Brookshire DS, Eubanks DS, Randall A. Estimating option price and existence values for wildlife resources. Land Economics 1983;59:1–15.

Brown TC, Gregory R. Why the WTA–WTP disparity matters. Ecological Economics 1999;28:323–35.

Brown TC, Champ PA, Bishop RC, McCollum DW. Which response format reveals the truth about donations to a public good? Land Economics 1996;72:152–66.

Cameron TA, Quiggin J. Estimation using contingent valuation data from a 'dichotomous choice with follow-up' questionnaire. Journal of Environmental Economics and Management 1994;27:218–34.

Cameron TA, Poe GL, Ethier RG, Schulze WD. Alternative non-market value-elicitation methods: are

the underlying preferences the same? Journal of Environmental Economics and Management 2002;44:391–425.

Carson RT. Three essays on contingent valuation (Welfare economics, non-market goods, water quality). PhD Thesis. Berkeley: Department of Agricultural and Resource Economics, University of California; 1985.

Carson RT, Steinberg D. Experimental design for discrete choice voter preference surveys. 1989 Proceeding of the Survey Methodology Section of the American Statistical Association. Washington, DC: American Statistical Association; 1990.

Carson RT, Flores NE, Martin KM, Wright JL. Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods. Land Economics 1996;72:80–99.

Carson RT, Flores NE, Meade NF. Contingent valuation: controversies and evidence. Environmental and Resource Economics 2001;19:173–210.

Choe KA, Whittington D, Lauria DT. The economic benefits of surface water quality improvements in developing countries: a case study of Davao, Philippines. Land Economics 1996;72:107–26.

Ciriacy-Wantrup SV. Capital returns from soil conservation practices. Journal of Farm Economics 1947;29:1181–96.

Coase R. The problem of social coast. Journal of Law and Economics 1960;3:1–44.

Coursey DL, Hovis JL, Schulze WD. The disparity between willingness to accept and willingness to pay measures of value. The Quarterly Journal of Economics 1987;CII:679–90.

Corso PS, Hammitt JK, Graham JD. Valuing mortality-risk reduction: using visual aids to improve the validity of contingent valuation. The Journal of Risk and Uncertainty 2001;23:165–84.

Cummings RG, Brookshire DS, Schulze WD, editors. Valuing environmental goods: a state of the arts assessment of the contingent valuation method. Totowa, NJ: Roweman and Allanheld; 1986.

Davis R. The value of outdoor recreation: an economic study of the marine woods. PhD Thesis. Harvard University; 1963.

Desvousges WH, Johnson FR, Dunford RW, Boyle KJ, Hudson SP, Wilson N. Measuring natural resource damages with contingent valuation: tests of validity and reliability. In: Hausman JA, editor. Contingent valuation: a critical assessment. Amsterdam: North Holland; 1993. p. 91–159.

Diamond PA, Hausman JA. On contingent valuation measurement of non-use values. In: Hausman JA, editor. Contingent valuation: a critical assessment. Amsterdam: North Holland; 1993. p. 3–38.

Diamond PA, Hausman JA. Contingent valuation: is some number better than no number? Journal of Economic Perspectives 1994;8:45–64.

Diamond PA, Hausman JA, Leonard GK, Denning MA. Does contingent valuation measure preferences? Experimental evidence. In: Hausman JA, editor. Contingent valuation: a critical assessment. Amsterdam: North Holland; 1993. p. 41–85.

Downing M, Ozuna Jr T. Testing the reliability of the benefit function transfer approach. Journal of Environmental Economics and Management 1996;30:316–22.

Duffield JW, Paterson DA. Inference and optimal design for a welfare measure in dichotomous choice contingent valuation. Land Economics 1991;67:225–39.

Foster V, Bateman IJ, Harley D. Real and hypothetical willingness to pay for environmental preservation: a non-experimental comparison. Journal of Agricultural Economics 1997;48:123–38.

Freeman III MA. The measurement of environmental and resource values. Washington, DC: Resource for the Future; 1993.

Griffin CC, Briscoe J, Singh B, Ramasubban R, Bhatia R. Contingent valuation and actual behavior: predicting connections to new water systems in the state of Kerala, India. The World Bank Economic Review 1995;9:373–95.

Hammitt JK, Graham JD. Willingness to pay for health protection: inadequate sensitivity to probability? Journal of Risk and Uncertainty 1999;18:33–62.

Hanemann MW. Welfare evaluations in contingent valuation experiments with discrete responses. American Journal of Agricultural Economics 1984;66:332–41.

Hanemann MW. Some issues in continuous and discrete response contingent valuation studies. Northeastern Journal of Agricultural and Resource Economics 1985;14:5–13.

Hanemann MW. Willingness to pay and willingness to accept: how much can they differ? American Economic Review 1991;81:635–47.

Hanemann MW. Valuing the environment through contingent valuation. Journal of Economic Perspectives 1994;8:19–43.

Hanemann MW, Kanninen B. The statistical analysis of discrete-response CV data. In: Bateman IJ, Willis KG, editors. Valuing environmental preferences. Oxford: Oxford Univ. Press; 1999. p. 302–441.

Hanemann MW, Loomis JB, Kanninen BJ. Statistical efficiency of double-bounded dichotomous choice contingent valuation. American Journal of Agricultural Economics 1991;73:1255–63.

Hanley N, Shogren JF, White B. Environmental economics in theory and practice. England: Macmillan; 1997.

Hansen TB. The willingness-to pay for the Royal Theatre in Copenhagen as a public good. Journal of Cultural Economics 1997;21:1–28.

Harrison GW. Valuing public goods with the contingent valuation method: a critique of Kahneman and Knestch. Journal of Environmental Economics and Management 1992;23:248–57.

Hausman JA, editor. Contingent valuation: a critical assessment. Amsterdam: North Holland; 1993.

Hicks J. Value and capital. Oxford: Oxford Univ. Press; 1946.

Hoehn JP, Randall A. Too many proposals pass the benefit cost test. American Economic Review 1989;79:544–51.

Hoen JP, Randall A. A satisfactory benefit cost indictor from contingent valuation. Journal of Environmental Economics and Management 1987;14:57–70.

Holms TP, Kramer RA. An independent sample test of yea-saying and starting point bias in dichotomous-choice contingent valuation. Journal of Environmental Economics and Management 1995;29:121–32.

Horowitz JK, McConnell KE. A review of WTA/WTP studies. Journal of Environmental Economics and Management 2002;44:426–47.

John KH, Walsh RG, Moore CG. Comparison of alternative nonmarket valuation methods for an economic assessment of a public program. Ecological Economics 1992;5:179–96.

Johannesson M, Jonsson B. Willingness to pay for antihypertensive therapy-results of a Swedish pilot study. Journal of Health Economics 1991;10:461–74.

Johannesson M, Johansson P-O, O'Conor R. The value of private safety versus the value of public safety. Journal of Risk and Uncertainty 1996;13:263–75.

Jones-Lee MW, Loomes MG, Philips P. Valuing the prevention of non-fatal road injuries: contingent valuation versus standard gambles. Oxford Economic Papers 1995;47:676–95.

Jordan JL, Elnagheep AH. Consequences of using different question formats in contingent valuation: a Monte Carlo study. Land Economics 1994;70:97–110.

Kahneman D. Comments on the contingent valuation method. In: Cummings RG, Brookshire DS, Schulze WD, editors. Valuing environmental goods: a state of the arts assessment of the contingent valuation method. Totowa, NJ: Roweman and Allanheld; 1986. p.185–94.

Kahneman D, Knestch JL. Valuing public goods: the purchase of moral satisfaction. Journal of Environmental Economics and Management 1992;22:57–70.

Kahneman D, Tversky A. On the psychology of prediction. Psychology Review 1973;80:237–51.

Kahneman D, Tversky A. Prospect theory: an analysis of decisions under risk. Econometrica 1979;47:263–91.

Kanninen BJ. Optimal experimental design for double-bounded dichotomous choice contingent valuation. Land Economics 1993;69:138–46.

Kealy MJ, Turner RW. A test of the equality of closed-ended and open-ended contingent valuations. American Journal of Agricultural Economics 1993;75:321–31.

Kealy MJ, Montgomery M, Dovidio JF. Reliability and predictive validity of contingent values: does the nature of the good matter? Journal of Environmental Economics and Management 1990;19: 244–63.

Kirchhoff S, Colby BG, LaFrance JT. Evaluating performance of benefit transfer: an empirical enquiry. Journal of Environmental Economics and Management 1997;33:75–93.

Knestch JL, Sinden JA. Willingness to pay and compensation demanded: experimental evidence of an unexpected disparity in measures of values. Quarterly Journal of Economics 1984;99:507–21.

Kristrom B, Laitila T. Stated preference methods for environmental valuation: a critical look (Paper submitted to). In: Folmer H, Tietenberg T, editors. International yearbook of environmental and resource economics. Cheltenham, UK: Edward Elgar; forthcoming.

Loomis JB. Comparative reliability of the dichotomous choice and open-ended contingent valuation techniques. Journal of Environmental Economics and Management 1990;18:78–85.

Loomis JB, duVair PH. Evaluating the effects of alternative risk communication devices on willingness to pay: results from a dichotomous choice contingent valuation experiment. Land Economics 1993;69:287–98.

Loomis JB, Gonzalez-Caban A, Gregory R. Do remainders of substitutes and budget constraints influence contingent valuation estimates? Land Economics 1994;70:499–506.

Louviere JJ. Conjoint analysis modeling of stated preferences. A review of theory, methods, recent developments and external validity. Journal of Transport Economics and Policy 1988;93–119 [January].

Louviere JJ, Woodworth G. Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. Journal of Marketing Research 1983;20:350–67.

Louviere JJ, Hensher DA, Swait J. Stated choice methods: analysis and application. Cambridge, UK: Cambridge Univ. Press; 2000.

McFadden D. Contingent valuation and social choice. American Journal of Agricultural Economics 1994;76:689–708.

Merrett S. Deconstructing households' willingness-to-pay for water in low-income countries. Water Policy 2002;4:157–72.

Mitchell RC, Carson RT. A contingent valuation estimate of national freshwater benefits: technical report to the U.S. Environmental Protection Agency. Washington, DC: Resources for the Future; 1984.

Mitchell RC, Carson RT. Using surveys to value public goods: the contingent valuation method. Washington, DC: Resource for the Future; 1989.

National Oceanic and Atmospheric Administration. Report of the NOAA Panel on contingent valuation. Federal Register 1993;58:4602–14.

Neill HR. The context for substitutes in CVM studies: some empirical observations. Journal of Environmental Economics and Management 1995;29:393–7.

Neill HR, Cummings RG, Gandeton PT, Harrison GW, McGuckin T. Hypothetical surveys and real economic commitments. Land Economics 1994;70:145–54.

Nickerson CA. Valuing public goods: a comment on Harrison's critique of Kahneman and Knestch. Journal of Environmental Economics and Management 1993;25:93–102.

Niklitschek M, Leon J. Combining intended demand and yes/no responses in the estimation of contingent valuation models. Journal of Environmental Economics and Management 1996;31:387–402.

Portney PR. The contingent valuation debate: why economists should care. Journal of Economic Perspectives 1994;8:3–17.

Randall A, Hoehn JP. Embedding in market demand system. Journal of Environmental Economics and Management 1996;30:369–80.

Randall A, Stoll JR. Consumer's surplus in commodity space. American Economic Review 1980; 70:449–57.

Randall A, Ives B, Eastman C. Bidding games for valuation of aesthetic environmental improvements. Journal of Environmental Economics and Management 1974;1:132–49.

Ready RC, Buzby JC, Hu D. Differences between continuous and discrete contingent value estimates. Land Economics 1996;72:397–411.

Samples KC, Hollyer JR. Contingent valuation of wildlife resources in the presence of substitutes and complements. In: Johnson RL, Johnson GV, editors. Economic valuation of natural resources: issues, theory and application. Boulder: Westview Press; 1990. p. 177–92.

Samuelson P. The pure theory of public expenditure. Review of Economics and Statistics 1954;36: 332–8.

Schulze WD, d'Arge RC, Brookshire DS. Valuing environmental commodities: some recent experiments. Land Economics 1981;57:151–69.

Schulze WD, McClelland G, Waldman D, Lazo J. Sources of bias in contingent valuation. In: Bjornstad DJ, Kahn JR, editors. The contingent valuation of environmental resources: methodological issues and research needs. Brookfield: Edward Elgar; 1996. p. 97–116.

Seip K, Strand J. Willingness to pay for environmental goods in Norway: a contingent valuation study with real payment. Environmental and Resource Economics 1992;2:91–106.

Shogren JF, Shin SY, Hayes DJ, Kliebenstein JB. Resolving differences in willingness to pay and willingness to accept. American Economic Review 1994;84:255–69.

Smith KV. Arbitrary values, good causes, and premature verdicts. Journal of Environmental Economics and Management 1992;22:71–89.

Smith KV. Nonmarket valuation of environmental resources: an interpretative appraisal. Land Economics 1993;69:1–26.

Smith VK, Osborne LL. Do contingent valuation estimates pass a 'scope' test? A meta-analysis. Journal of Environmental Economics and Management 1996;31:287–301.

Teisl MF, Boyle KJ, McCollum DW, Reiling SD. Test–retest reliability of contingent valuation with independent sample pretest and posttest control groups. American Journal of Agricultural Economics 1995:613–9.

Thaler R. Toward a positive theory of consumer choice. Journal of Economic Behavior and Organization 1980;1:39–60.

Thompson MS, Read JL, Lian M. Feasibility of willingness-to-pay measurement in chronic arthritis. Medical Decision Making 1984;4:195–215.

Thompson E, Berger M, Blomquist G, Allen S. Valuing the arts: a contingent valuation approach. Journal of Cultural Economics 2002;26:87–113.

Venkatachalam L. Economic valuation of water used in the household sector: a contingent valuation approach in a developing country context. Unpublished PhD Thesis. Chennai: University of Madras; 2000.

Walsh RG, Loomis JB, Gillman RA. Valuing option, existence and bequest demands for wilderness. Land Economics 1984;60:14–29.

Welsh MP, Poe GL. Elicitation effects in contingent valuation: comparisons to a multiple bounded discrete choice approach. Journal of Environmental Economics and Management 1998;36:170–85.

Whitehead JC, Blomquist GC. Measuring contingent values for wetlands: effects of information about related environmental goods. A Research Report; 1990.

Whittington D. Administering contingent valuation surveys in developing countries. World Development 1998;26:21–30.

Whittington D, Briscoe J, Mu X, Barron W. Estimating the willingness to pay for water services in developing countries: a case study of the contingent valuation in Southern Haiti. Economic Development and Cultural Change 1990;38:293–312.

Whittington D, Lauria DT, Mu X. A study of water vending and willingness to pay for water in Onitsha, Nigeria. World Development 1991;19:179–98.

Whittington D, Smith VK, Okorafor A, Okore A, Jin Liu L, McPhail A. Giving respondents time to think in contingent valuation studies: a developing country application. Journal of Environmental Economics and Management 1992;22:205–25.

Willig RD. Consumer's surplus without apology. American Economic Review 1976;66:589–97.

Dr. L. Venkatachalam is an Assistant Professor in Economics at the Institute for Social and Economic Change, Bangalore, India. Presently, he works on nonmarket valuation, natural resource accounting and economics of water.