Annex B

Contingent Valuation as a Means of Valuing the Conservation of Coral Reefs: An Assessment of the Method

Nick Hanley
Institute of Ecology and Resource Management, University of Edinburgh, Edinburgh, Scotland, United Kingdom

The contingent valuation method (CVM) is a means of assigning monetary values to resources and service flows that are unpriced or under-priced by the market. CVM is based on neo-classical welfare economics, where the value of an environmental resource to an individual is expressed either as their maximum willingness-to-pay (WTP) to acquire or safeguard it, or else the minimum monetary compensation they would accept to go without an increase in that good or tolerate a decrease (willingness-to-accept compensation; WTAC). Given that it is "missing markets" that result in such environmental resources as clean air, coral reefs or biodiversity being unpriced, CVM relies on a constructed, hypothetical market to produce monetary estimates of value. The researcher obtains peoples' bids (either WTP or WTAC) for a specified change in the environmental good of interest *contingent* on the description of this hypothetical market. Thus, an individual's WTP or WTAC can be expected to depend on:

- The description of the contingent market;
- What they know about the environmental good, which depends partly on what they are told about it as part of the CVM survey;
- Their own preferences;
- Their budget constraints; and,
- The availability of substitutes and complements.

Empirically, it also turns out that stated WTP (throughout the rest of this annex, WTP will be used to refer to both itself and WTAC unless otherwise stated) also depends on the design of the constructed market and how responses are subsequently analyzed.

Historically, CVM developed through gradual acceptance and use by United States government agencies. An important milestone was the acceptance by the United States courts of the use of CVM in natural resource damage assessments under the 1980 Comprehensive Environmental Response, Compensation and Liability Act (CERCLA). However, perhaps the most relevant event in the development of CVM was the case pursued by the State of Alaska and the federal government in the United States against Exxon as a result of the Exxon Valdez oil spill (Bateman and Willis 1999). This led to the establishment of the so-called Blue Ribbon Panel, out of which emerged National Oceanic and Atmospheric Administration (NOAA) guidelines on the use of CVM, especially regarding non-use values (Federal Register 1993, 1994). These guidelines are still the subject of some debate, but form the basis for the design elements presented in this annex.

Essentially, a CVM exercise consists of: i) describing the environmental change in question; ii) describing the contingent market; iii) establishing a bid vehicle and reason for payment; iv) seeking bids, either through an open-ended format, a bidding game, a payment card, or a single or double-bounded dichotomous choice mechanism; v) estimating mean or median WTP; iv) aggregating this average bid to a population total value; and, vii) carrying out reliability and validity tests of the CVM results.

Principal Problem Areas

The CVM method has been used to estimate the value of a wide variety of environmental resources, including air and water quality, outdoor recreation, and landscape and wildlife conservation. These applications have highlighted some general problems in CVM—namely, the concept of utilitarian values for environmental resources (Spash and Hanley 1995; Stevens *et al*. 1991); benefits transfer (Bergland *et al*. 1995); nesting and part-whole bias (Boyle *et al*. 1994); calibration and aggregation (Foster *et al*. 1998); and the concept of non-use values (Bishop and Welsh 1992). In addition, the ability of researchers using

CVM to value the different characteristics that make up, say, a pleasant landscape, has been limited (e.g., Hanley and Ruffell 1993). This has turned attention to other stated preference approaches, notably choice experiments (Adamowicz *et al*. 1994).

Early work (summarized in Mitchell and Carson 1989) tended to concentrate on what might be termed "design bias" effects; these included the impact on WTP of changes in the starting point in bidding games and tests for strategic behavior. Later, much attention was, in contrast, given to large differences between WTP and WTAC measures of value, which were inconsistent with mainstream welfare economics predictions. These differences have now been attributed to endowment effects (Knetsch 1989) and/or to substitution effects (Hanemann 1991). Another recent trend has been the large number of articles concerned with optimal design and subsequent econometric analysis of dichotomous choice models. Other papers (e.g., Munro and Hanley 1999) have shown that changes in the information set provided to respondents in a CVM survey can significantly affect their WTP, but that this is a desirable characteristic of the method.

Perhaps the four main current worries within CVM are: i) part-whole bias; ii) lexicographic preferences; iii) non-use values; and, iv) aggregation. These problems are now briefly described, before tentative best practice guidelines are outlined.

Nesting and Part-Whole Effects

It is well known that in CVM a good will be valued higher when valued in isolation than when as part of a more inclusive bundle. This is to be expected if goods within the bundle are substitutes for each other, to a degree, in terms of the utility they generate (Carson *et al*. 1998). This phenomenon has variously become known in CVM as embedding, nesting and part-whole bias, and may well exist for private market goods as well as for non-market goods. One possible "cure" for this problem is to ask respondents to bid for the more inclusive good first, and then to apportion some amount of this total bid to the good being valued.

Willis *et al*. (1993), for example, used this approach in their study of English environmentally sensitive areas (ESAs). Respondents were first asked to state a WTP amount, in terms of additional income tax, to maintain the entire ESA program in England and Wales. Residents and visitors were then shown pictorial and textual representations of what the landscape in either ESA would look like with and without the ESA scheme in place. Respondents stated which landscape they preferred and were then asked to allocate an amount for that landscape from their already declared ESA "budget", having been told that money "spent" on one ESA could not be spent on another (in other words, that there was an opportunity cost).

From the CVM literature and from economic theory, we know that this procedure will elicit lower bids for an individual ESA than when that ESA is bid for alone. This procedure might be seen as desirable in the sense that it presents a direct opportunity cost for bidding for any individual ESA (less can be "spent" on the other nine) and, also, that it produces more conservative value estimates. However, this procedure suffers from one major problem. As has been noted above, respondents who are not familiar with the good being valued must be given enough accurate and unbiased information to permit them to make well informed choices. ESAs are, for most individuals, unfamiliar goods in terms of the benefits they generate. Providing "full" information on each ESA in this instance would be an impractical task. Thus, this method of dealing with nesting effects is flawed. Additionally, one might ask why respondents should not first be asked to allocate some total for all public environmental spending, then allocate some of this to the ESA program, and then allocate some of this to a specific ESA. But why stop there? On the same logic, respondents should surely first be asked about how much they are willing to pay in taxes for total government expenditure. Yet this seems beyond the original intention or capabilities of CVM, especially when one considers the information issue.

Lexicographic Preferences

Both Stevens *et al*. (1991) and Spash and Hanley (1995) have found evidence that when people are asked to participate in CVM surveys concerned with wildlife protection, a proportion of these respondents have preferences that are at odds with the utilitarian ethic and the demand model underlying cost-benefit analysis (CBA). In essence, such individuals (approximately 25% of the sample in each case) refuse the concept of trading off income changes for changes in the level of environmental quality. Spash and Hanley (1995) argue that such preferences may be characterized as "lexicographic", derived from an ethical system based on rights. The implication is that WTAC amounts for such individuals will be infinite, and WTP amounts will be either zero (i.e., the individual protests) or a positive amount that does not vary with the level of environmental change involved. Since the behavior of such individuals does not correspond to the model underlying CBA, they are effectively disenfranchised by the CBA process. Identifying such individuals is clearly

important, although what to do about them is much less clear. It also seems important to test, empirically, what determines such behavior and whether it is independent of the opportunity cost of, in this case, wildlife protection. The issue of lexicographic preferences in particular, and non-utilitarian ethics in general, within CBA is currently unresolved.

Non-Use Values and Obscure Resources

Non-use (passive use) values have long been a subject of some controversy in contingent valuation. Non-use values represent the utility derived from individuals from the existence of an environmental resource when they do not consume it *in situ* (e.g., by bird-watching or hunting). Arguments for and against the acceptability of non-use values can be found, for example, in Randall (1993) and Hausman (1993). An interesting finding in the CVM literature is that non-use values appear to exist for respondents who were not aware of the good before the survey took place. An example of this phenomenon is reported by Bishop and Welsh (1992) who note that the citizens of Wisconsin are apparently willing to pay US$12 million for preserving the striped shiner, a "...small minnow inhabiting the turbid depths of the Milwaukee River" of which few respondents were aware prior to the survey. Bishop and Welsh (1992, p.138) contend that these values are as real as non-use values for well-known resources such as the Grand Canyon: "...lack of knowledge cannot be taken as evidence that the existence of such resources lacks the ability to satisfy preferences…It could simply indicate the lack of past choice opportunities to motivate information gathering. In the case of the striped shiner, it is possible that people are concerned about the fate of endangered species, even obscure ones."

Thus, lack of *ex ante* knowledge is not a reason for a non-credible WTP value, especially as we have already argued that the CVM process is an information providing process that is *expected* to change preferences. Whether values for those in the sample who were ignorant of the resource prior to the survey can be used to say anything about the values of those outside the sample who have not so been informed is, however, a moot point.

Aggregation Problems

Some of the problems of aggregating benefit estimates in CVM studies are largely of a practical nature (e.g., estimating total visitors to a wilderness area). Estimates can, of course, be made. However, with regard to projects where the general public can be expected to benefit, two awkward questions arise. First, which population do we count as valid? Multiplying even very small per person values by national populations give rise to very large aggregate non-use values. Second, are the large aggregate values that arise credible? Bishop and Welsh (1992) refer to an "adding up" problem for non-use values whereby, possibly due to their symbolic value, individuals would give identical WTP values for *any* environmental good cause that they are made aware of, but their WTP for all of these projects added together would not be equal to the sum of these individual amounts.

For this reason, and also because of a worry that the very hypothetical nature of the CVM situation causes an inflation of stated values, economists have suggested "calibrating" CVM estimates when aggregated into smaller amounts (Foster *et al*. 1998; NOAA 1993). Foster *et al*. (1998) report ratios of stated to actual WTP for wildlife conservation in the United Kingdom in the range of 0.3 to 10.5 (with a mean of 2.93), although they also note that the methods adopted in the studies from which these numbers arise vary widely, making comparisons difficult. NOAA (1993) suggest a somewhat ad hoc calibration figure of 50% in the absence of an experimental study undertaken alongside any given CVM study, yet this neglects the probable range of calibration desirable in different contingent markets across the non-use/use and public/private good continuums.

Design Elements—A Best Practice Guide

In discussing current views on what constitutes "best" design in CVM, it is first necessary to describe the nature and evolution of United States government guidelines on the use of the technique, which seem likely to heavily influence the acceptability of CVM results in the United States. The wrecking of the oil tanker the Exxon Valdez off the coast of Alaska in 1989 was the somewhat unforeseen cause of a major spur to the development of CVM in terms of a legally acceptable method of valuing environmental damages in the United States. United States law had gradually seen the introduction of damage claims for environmental losses, principally under the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) regulations of 1986 and the Oil Pollution Act of 1990. Following a famous judgment by the DC Court of Appeals (State of Ohio versus Department of the Interior), non-use (or more strictly, what has been termed "passive use" values, including the values derived from watching wildlife on TV, for example) were deemed relevant under this body of legislation in that persons could sue responsible parties for lost passive

use values. This clearly had an enormous implication for Exxon since many of the environmental damages resulting from the spill (i.e., damage to wildlife and a pristine, fragile ecosystem) were likely to be passive use values, as opposed to actual active use values, since actual active use of the area was relatively modest.

As a counter to the possibly large size of damage claims being made against Exxon, the company funded a series of studies that basically tried to discredit CVM as a method for valuing losses in passive use values (Cambridge Economics 1992). The government body responsible for issuing regulations on the assessment of damages from oil spills, the National Oceanic and Atmospheric Administration (NOAA), convened a panel of distinguished economists thought to have no vested interest in the CVM method to conduct hearings on the validity of the method in 1992. Members of the panel were Robert Solow, Kenneth Arrow, Edward Leamer, Paul Portney, Roy Radnor and Howard Schuman. The panel's report on their findings was published in January 1993 (Federal Register 15.1.93) and was basically a cautious acceptance of CVM for valuing environmental damages, including lost passive use values. These findings have recently been developed as a set of proposed guidelines for future legally admissible CVM studies, which seem bound to at least influence the future development of the method (Federal register 7.1.94). The principle recommendations were as follows:

1. A dichotomous choice format should be used;
2. A minimum response rate from the target sample of 70% should be achieved;
3. In-person interviews should be employed (not mail shots) with some role for telephone interviews in the piloting stages;
4. WTP, not WTAC, measures should be sought;
5. After excluding protest bids, a test should be made of whether WTP is sensitive to the level of environmental damage;
6. CVM results should be calibrated against experimental findings, otherwise a 50% calibration factor should be applied to CVM results;
7. Respondents should be reminded of their budget constraints; and,
8. Respondents should be given "adequate" information about the environmental change in question.

These measures are, at the very least, a rather strange mixture of theoretically based recommendation and crude "rules of thumb". Taken together, they make contingent valuation a very expensive exercise if implemented in full. It would be unfortunate if all CVM practitioners felt constrained to stick to these guidelines in future research, since the guidelines pose some awkward questions. These include:

- Are all "protest" bids giving the same signals and how should these signals be interpreted and utilized in any case?
- Can the 50% calibration factor be justified empirically?
- How can the weaknesses of the dichotomous choice design format be overcome?

How Should a CVM be Designed?

Credibility of Hypothetical Market

It is essential that the hypothetical market used be credible and, if possible, rely on routine or previously experienced behavior on the part of respondents. Such credibility can be tested for in focus groups (see below). Credibility as a concern extends to the bid vehicle used (thus respondents must be able to envisage that it could be collected); the bid vehicle should also be, as far as possible, uncontroversial.

Protest Bids

Protest bids (i.e., zero bids for reasons other than a true zero value being placed on the environmental change) should be identified, and reasons for them sought. CVM surveys that suffer from very high levels of protesting (e.g., more than 40% of all respondents) might be challenged as either having used a non-credible hypothetical market, having used a controversial bid vehicle, or having involved a radical change in implied property rights.

Information

Individuals need to be informed about all important aspects of the resource concerned and the nature of the change being considered. However, this information needs to be provided in a manner which ordinary people can comprehend (testing for comprehension is another function of focus groups). There is clearly a trade off between the amount of information provided and the extent to which people can assimilate and understand it within the normal time-span of a CVM survey. No firm guidelines can be provided here. Rather, the researcher must reach a common sense compromise. Focus groups can identify which aspects of the resource or resource change are deemed as most important by the individuals concerned.

Careful Survey Instrument Development

The key feature of any CVM study is the questionnaire itself. A successful questionnaire design is now recognized to involve three primary steps: i) use of focus groups to find out how respondents identify with the resource in question, what language they use to describe it, and their understanding of draft survey materials; ii) use of verbal protocols, whereby respondents complete draft questionnaires in a "thinking aloud" mode, to enable researchers to understand how people will react to survey questions and how they will form their answers; and, iii) pilot surveys to pre-test aspects such as design of dichotomous choice bid levels.

Choice of Bid Collection Technique

As was mentioned above, CVM researchers can use open-ended, bidding game, payment card, or discrete choice (referendum type) designs. The question of which is preferable is still largely unresolved. Bidding games often suffer from starting point bias, while payment cards suffer from anchoring effects. Open-ended designs may be more difficult for respondents to complete and may encourage more strategic behavior. Since the NOAA study recommends dichotomous choice approaches, much recent attention has been focused on this method.

In a closed-ended referendum, a single payment is suggested to which respondents either agree or disagree (i.e., a yes or no reply). The calculation of mean or median WTP from such responses is more complex than the alternatives above, since all that is revealed to the researcher is whether the respondent is willing to pay a particular sum (known as the offer price). The researcher must then either make assumptions about the underlying distribution of true WTP, or else use non-parametric techniques. Double-bounded referendum models present those respondents who say "no" to the first amount with a lower amount and those respondents who say "yes" to the first amount with a higher amount, thus eliciting increased information (e.g., Carson *et al*. 1994). Finally, uncertainty over valuation can be allowed for in both open and closed-ended valuation methods (e.g., Ready *et al*. 1999).

Discrete choice approaches have consistently produced higher estimates of value than open-ended approaches due to the phenomena of "yea-saying" and preference uncertainty. In addition, they require larger sample sizes, while the calculation of mean WTP is influenced by the functional form of the logit equation, the extent of any truncation used, and the design of bid levels (number and amounts). Finally, there is some evidence that discrete choice approaches suffer from higher hypothetical market bias than open-ended approaches, although strategic behavior is likely to be greater in open-ended formats, which may also result in a greater degree of non-response.

Means of Sample Collection

CVM responses may be collected by mail shot, face-to-face surveys or telephone surveys. Of these three alternatives, telephone surveys are usually considered least desirable. Face-to-face surveys are recommended by the NOAA panel, but the context in which the survey is conducted is important (e.g., shopping malls versus people's homes). Such surveys are also relatively expensive. Mail shots are prone to low response rates and non-response bias, and the order in which the questions are answered is hard to control. However, mail shots can be very cost effective. Some surveys comparing mail shot with face-to-face questioning have found no significant differences in WTP, so long as high response rates can be achieved (i.e., greater than 40%).

Tests of Sensitivity to Scope

It is important to show that WTP is sensitive to the scale of the environmental change involved, where "scale" is defined in accordance with respondents' perceptions. Thus, WTP to protect one coral reef should be less than WTP to protect all coral reefs, although marginal WTP is expected to decline. The NOAA guidelines also recommend such tests of scope.

Tests for Reliability and Validity

Tests for reliability and validity should be incorporated in every CVM. These tests will involve some or all of the following:

- *Convergent validity test*. Does CVM produce similar results for a given resource change as alternative valuation techniques?
- *Theoretical construct validity test*. Can WTP responses be explained statistically to a satisfactory level and in a way in accordance with theoretical expectations? For example, does WTP increase with income? The alternative hypotheses here are that WTP measures are random numbers and/or are not in accord with economic theory.
- *Test-retest criterion*. If the CVM survey is repeated on a different sample drawn from the same population, do statistically different results emerge?
- *Calibration*. Can CVM results be related to actual payments for the resource in question?

- *Debriefing*. Have respondents understood the questions asked of them? Have they valued the same change in resource allocation that the researcher wished?

Special Features of CVM in Developing Country Applications

Many applications of CVM now exist in developing countries (for a survey, see Wasike 1996). These may be seen to have resulted from both academic interest in whether CVM could be transferred to a developing country context, and from policy and project needs on the part of agencies such as the World Bank and the United States Agency for International Development (USAID). Examples of developing country applications of CVM include Adger *et al*. (1994) on the value of Mexican forests, Navrud and Mungatana (1994) on the recreational value of wildlife viewing in Kenya, and Swallow and Wouldyalew (1994) on tsetse fly control in Ethiopia. In addition, many authors have applied CVM in developing countries to issues of drinking water supply. These include Briscoe (1990) for Brazil, Whittington *et al*. (1990) for Haiti, and McPhail (1994) for Tunisia.

Issues emerging from these applications include:

- *Low income levels on the part of respondents*. Income is often strongly related to WTP in these surveys, yet income levels are often low. Some authors have experimented with WTP denominated in units other than money. Swallow and Woudyalew (1994) used willingness to contribute labor hours, while Shyamsundar and Kramer (1996) used rice as a means of payment.
- *Irregularity of income flows*. In subsistence and other types of farming, income flows may be very irregular. Combined with imperfect markets for credit, this means that the temporal nature of payments in a CVM may be important (Wasike and Hanley 1998).
- *Contextual impacts*. These include the presence of listeners during surveys and "prestige" effects. Whittington *et al*. (1993) found statistically significant effects on WTP of such contextual factors in Ghana.
- *Cultural views on environmental values*. The western notion of CBA as a means of taking decisions, and of individual preferences as the measure of environmental values, may fit poorly with prevailing cultural views and values.

However, the consensus emerging from the growing number of CVM applications in developing countries is that, provided the analysis is sufficiently attuned to local circumstances, the method can be successfully applied.