

# Testing and Calibrating the Measurement of Nonmarket Values for Oil Spills via the Contingent Valuation Method

**Final Technical Summary** 

**Final Study Report** 



U.S. Department of the Interior Minerals Management Service Pacific OCS Region

# Testing and Calibrating the Measurement of Nonmarket Values for Oil Spills via the Contingent Valuation Method

## **Final Technical Summary**

## **Final Study Report**

Authors

W. Michael Hanemann Jon A. Krosnick Principal Investigators

Prepared under MMS Cooperative Agreement No. 14-35-0001-30758 by Coastal Marine Institute Marine Science Institute University of California Santa Barbara, CA 93106

U.S. Department of the Interior Minerals Management Service Pacific OCS Region

Camarillo February 2006

#### Disclaimer

This report has been reviewed by the Pacific Outer Continental Shelf Region, Minerals Management Service, U.S. Department of the Interior and approved for publication. The opinions, findings, conclusions, or recommendations in this report are those of the authors, and do not necessarily reflect the views and policies of the Minerals Management Service. Mention of trade names or commercial products does not constitute an endorsement or recommendation for use. This report has not been edited for conformity with Minerals Management Service editorial standards.

#### **Availability of Report**

Extra copies of the report may be obtained from:

U.S. Dept. of the Interior Minerals Management Service Pacific OCS Region 770 Paseo Camarillo Camarillo, CA 93010 Phone: 805-389-7621

A PDF file of this report is available at: http://www.coastalresearchcenter.ucsb.edu/CMI/

#### **Suggested Citation**

The suggested citation for this report is:

Hanemann, W. M., and J. Krosnick. *Testing and Calibrating the Measurement of Nonmarket Values for Oil Spills via the Contingent Method*. MMS OCS Study 2006-010. Coastal Research Center, Marine Science Institute, University of California, Santa Barbara, California. MMS Cooperative Agreement Number 14-35-0001-30758. 27 pages.

## **Table of Contents**

FINAL TECHNICAL SUMMARY	1
FINAL STUDY REPORT	5
Review of Previous Studies	5
Overview of Our Investigation	7
Results	9
Reanalysis of Prior Survey Studies	9
Experiments with Student Participants	11
Study 2	12
Simulation	15
National Survey	15
Conclusion	18
References	19

## List of Tables

<b>Table 1.</b> Reanalysis of Green at al. Data	. 10
Table 2. Reanalysis of Montrose Data	. 11
Table 3. Reanalysis of California Oil Spill Data	. 11
Table 4. Mean "Yes" Responses at Each Percentile Bid by Question Type	. 13
<b>Table 4a.</b> "Yes" Responses from Greater than + Less than Format = 1	. 13
<b>Table 5.</b> Difference between Percentage of "Yes" Responses and the True	
"Percentiles"	. 14

## FINAL TECHNICAL SUMMARY

**STUDY TITLE**: Testing and Calibrating the Measurement of Nonmarket Values for Oil Spills via the Contingent Valuation Method.

**REPORT TITLE**: Testing and Calibrating the Measurement of Nonmarket Values for Oil Spills via the Contingent Valuation Method.

**CONTRACT NUMBER**: 14-35-0001-30758

SPONSORING OCS REGION: Pacific

APPLICABLE PLANNING AREA(S): Southern California

FISCAL YEAR(S) OF PROJECT FUNDING: FY 95

COMPLETION DATE OF REPORT: February, 2006

COST(S): FY 95: \$103,867; FY 96: no-cost

CUMULATIVE PROJECT COST: \$103,867

PROJECT MANAGER: Russell J. Schmitt

**ADDRESS**: Coastal Research Center, Marine Science Institute, University of California, Santa Barbara, CA 93106-6150

PRINCIPAL INVESTIGATORS: W. Michael Hanemann<sup>1</sup> and Jon A. Krosnick<sup>2</sup>

**ADDRESSES:** <sup>1</sup> Department of Agriculture and Resource Economics, 327 Giannini Hall, University of California Berkeley, CA 94710, <sup>2</sup> 450 Serra Mall, Stanford University, Stanford, CA 94305

**KEY WORDS**: nonmarket values, Contingent Valuation (CV) method, WTP values, oil spills, open-ended questions, closed-ended questions, willingness-to-pay

## **BACKGROUND**:

Data on economic valuation are needed inputs in making policy and management decisions for coastal and ocean resources. Some of the economic aspects involve commercial activities such as fishing. Others involve what are known as non-market benefits, such as private recreation and aesthetics. For recreation and what are called *use values*, an economic technique known as the *travel cost (TC) method* can be used. For other aspects - aesthetics, concern for wildlife and what are known as *non-use values* - the approach used is to interview people directly and elicit from them an estimate of what they would be willing to pay (WTP) to prevent damage to these resources. This is known as the *contingent valuation* (CV) method. The 1990 Oil Spill Act

(OPA) calls for measurement of non-use values in analyzing the damages from oil spills, and CV has come to be seen as the key way to collect this information.

The CV method was first proposed by Ciriacy-Wantrup (1947) as a means of non-market nonvaluation in the specific context of protection against soil erosion. Ciriacy-Wantrup argued that the appropriate analogy for this was not private choices as expressed through individual purchases of normal market commodities but rather collective choice through voting on the provision of a public good. He saw the CV survey as a surrogate for voting - one would approach citizens to see if the item were worth what it cost to provide. The first CV survey was conducted by Davis (1963) to value recreation in the Maine woods. A decade later, Randall et al (1974) conducted the first major non-use value CV study on air quality and visibility in the Four Corners area. By the late 1970's, CV studies were being commonly performed to evaluate environmental and other non-market commodities. The 1980's saw several important methodological developments, including collaboration between economists and other social scientists with expertise in survey research. Two landmarks were an EPA conference in Palo Alto in 1984 that brought together leading CV practitioners, other distinguished economists and psychologists to assess the then state-of-the-art [Cummings et. al (1986)] and the publication of what has become the standard reference on CV, Mitchell and Carson (1989), placing it in the broader context of economics, sociology, psychology, market research and political science. This was a collaboration between an environmental sociologist, Mitchell, and an environmental economist, Carson, who had received his Ph.D under Hanemann at U.C. Berkeley.

One of the changes in CV methodology that occurred during the 1980's was a shift in question format. The early CV studies had used an open-ended question along the lines of "What is the most that you would be willing to pay for the item?" The alternative is to use an open-ended format: "If this item cost you \$x, would you be willing to pay that much?" Different amounts, x, are presented to different respondents; their responses trace out a *bid-response function* that show the percent willing to pay an amount as a function of the amount. From this one can readily deduce the mean or median willingness to pay (WTP) in the population surveyed.

If there were a referendum on the item, as with propositions in the California Ballot, the median would correspond to passage with majority voting. Hence, the approach is also known as the *referendum* method. This was first used by Bishop and Heberlein (1979) in a study of duck hunting in Wisconsin. It was popularized by Hanemann (1984) who showed how the responses could be interpreted in terms of an economic model of utility maximization, and who developed the formulas to calculate estimates of mean or median WTP from survey responses. Since the late 1980's, it has become the standard approach for many CV studies. It has been extended in several ways by Hanemann, Loomis and Kanninen (1991) and Cooper and Hanemann (1994) to increase its statistical power. Evidence suggests that respondents generally find it easier to respond to the closed-ended valuation question, and often more meaningful - it is, indeed, like voting in a referendum which, even in rare in practice, seems like a natural thing. When NOAA created a Blue Ribbon Panel to advise it on the use of CV for measuring non-use values, the panel took this view and endorsed the referendum approach [Arrow et al. (1993)]. However, the Panel indicated the need for further research on the implementation and calibration of the referendum approach, and that is the focus of the current proposal.

The empirical application in the present study builds on existing CV research on California oil spills by Hanemann and Krosnick. In 1988, Hanemann was asked by the California Attorney General's Office to assist it in seeking natural resource damages following an oil spill at the Shell Oil refinery in Martinez, CA. This work led to a \$20 million settlement in 1989, the largest payment ever made in the US up to that time for a natural resource damages claim [Hanemann (1992)]. As part of the settlement, \$645,000 was set aside for a CV study to be directed by Hanemann aimed at valuing damages that might occur in a future oil spill in California (this sum was recently raised to \$795,000 to cover additional survey costs). State and federal officials involved in the Shell spill felt that having such information available on the shelf would be very helpful for future planning and policy analysis as well as damage assessment.

The study was conducted under a contract between the California Attorney General Office and Hanemann. It involved a team of economists, sociologists and psychologists at several universities. Jon Krosnick was a member of this team and was actively involved in developing the final survey instrument. The study was conducted as academic research, not for litigation, and the results are being published as a book. Economists with OSPR in the California Department of Fish & Games and with the NOAA Oil Spill Office in the US Department of Commerce served as peer reviewers for the California Attorney General Office, together with academic peer reviewers. The survey involved in-person interviews with a statewide sample of approximately 1,000 households conducted by a leading national survey company, Westat Inc, under sub-contract to Professor Hanemann.

During the course of developing the survey instrument, several prototypical oil spill scenarios were developed as candidates for valuation. In the end, it was found infeasible to include more than one of the scenarios in an interview while remaining within the time parameters specified in Westat's contract. The team decided to focus on one spill scenario, with the idea that OSPR or other agencies could subsequently replicate the survey with a different oil spill scenario if so desired. The basic structure of the survey instrument and the survey logistics would already have been developed and proven up in the current survey, thereby making subsequent replication both simpler and less costly. Thus, the survey was viewed as a cornerstone for future state and federal work on systematic valuation of the effects of oil spills in California.

Although the Shell settlement has funded an extensive survey, it provided no funds for methodology development or research on the effects of alternative survey designs. That is the focus of the current research. The aim was to investigate an important methodological issue which is currently the subject of much discussion among practitioners, namely the effects of question format and structure. The purposes of our work were (1) to determine how the results of the Shell-funded CV survey should be interpreted or adjusted in order to provide a robust estimate of WTP values for oil spills in California, and (2) to generate findings of benefit for CV researchers generally about how to conduct future studies using this method.

### **OBJECTIVES**:

The aim of this research was to extend and improve the contingent valuation (CV) method with particular reference to valuing California coastal and estuarine resources. The specific objective was to investigate how the results of CV surveys are affected by the form of the valuation question -- discrete-response ("referendum") versus continuous-response.

### STUDY PRODUCTS TO DATE:

#### **Publications:**

- Beatty, T. 2001. <u>Three Essays on Contingent Valuation</u>. Ph.D Dissertation, Department of Agricultural & Resource Economics, UC Berkeley.
- Holbrook, A.L., J.A. Krosnick, R.T. Carson, and R.C. Mitchell. 2000. Violating conversational conventions disrupts cognitive processing of attitude questions. *Journal of Experimental Social Psychology*, 36, 465-494.

#### **Presentations:**

- Hanemann, W.M., J. Cooper, and G. Signorello. 2000. One-and-One-Half Bound Dichotomous Choice Contingent Valuation. Invited paper presented at the Conference on the Theory and Application of Environmental Valuation, Kobe University, Japan.
- Hanemann, W.M., J.A. Krosnick, and T. Beatty. 2000. Closed vs. Open Ended Approaches to CV: Explaining Why results are Different. Paper presented at W133 Annual Research Meeting, Sheraton Kauai Resort, Hawaii.
- Hanemann, W.M., J.A. Krosnick, and T. Beatty. 2000. Closed vs Open Ended Approaches to CV: Explaining Why results are Different. Paper presented at Fifth Occasional Workshop on Environmental and Resource Economics, UC Santa Barbara.
- Holbrook, A.L., J.A. Krosnick, R.T. Carson, and R.C. Mitchell. 1999. Violating conversational conventions disrupts cognitive processing of attitude questions. Paper presented at the American Association for Public Opinion Research Annual Meeting, St. Petersburg, Florida.
- Holbrook, A., J.A. Krosnick, R.T. Carson, and R.C. Mitchell. 1998. Violating conversational conventions disrupts cognitive processing of survey questions. Paper presented at the American Association for Public Opinion Research Annual Meeting, St. Louis, Missouri.
- Shaeffer, E.M., S.F. Lampron, J.A. Krosnick, T.N. Tompson, P.S. Visser, and W.M. Hanemann. 2004. A comparison of open vs. closed survey questions for valuing environmental goods. Paper presented at the American Association for Public Opinion Research Annual Meeting, Phoenix, Arizona.

## FINAL STUDY REPORT

### **Review of Previous Studies**

In this paper, we explore whether open and closed questions do indeed yield comparable results, with a focus on a particular application: Contingent valuation (CV) surveys. Contingent valuation is an unusual survey methodology designed to achieve an unusual purpose in an unusual way (Mitchell and Carson 1989). These surveys are unusual mostly because they do not simply involve asking people questions about their opinions on matters of public discussion. Rather, CV questionnaires typically begin by presenting a large amount of information to respondents about a set of circumstances with which they are probably not familiar, sometimes lasting as long as 30 minutes. Then respondents about a particular program that could be implemented by a government agency and were asked a question to determine how much money the respondents would be willing to pay toward covering the cost of the program.

Some studies have asked this question in an open format, asking respondents what is the most they would be willing to pay. Other studies have asked this question in a closed format, inquiring about a respondent's willingness to pay a particular stated price. Different respondents were randomly assigned to be asked about different prices, and the aggregate pattern of responses could be used to yield a demand curve, from which the mean, median, and total maximum willingness to pay for the entire sample could be derived. A number of studies have compared the results obtained by closed-ended and open-ended questions measuring willingness to pay (WTP). And these studies have consistently found larger mean values are obtained by closed questions than by open ones (Balistreri, McClelland, & Schulze, 1994; Balistreri, McClelland, Poe, & Schulze, 1995; Bohara, McKee, Berrens, Jenkins-Smith, Silva, and Brookshire, 1998; Boyle, Johnson, McCollum, Desvousges, Dunford, & Hudson, 1996; Brown, Champ, Bishop, & McCollum, 1996; Casey & Delquie, 1995; Desvousges, Johnson, Dunford, Boyle, Hudson, & Wilson, 1993; Green, Jacowitz, Kahneman, & McFadden, 1998; Johnson, Bregenzer, & Shelby, 1990; Kealy & Turner, 1993; Kristrom, 1990, 1993; Loomis, Lockwood, & DeLacy, 1993; McFadden, 1994; McFadden & Leonard, 1993; Seller, Stoll, & Chavas, 1985; Walsh, Johnson, & McKean, 1992; cf. Frykblom, 1997). Some of this discrepancy is due to inappropriate calculation methods being applied to answers to the closed questions in order to generate sample means (Halvorsen & Soelensminde, 1998; Huang &

Smith, 1998). But even after correcting calculation methods, the difference between question forms remains apparent (Halvorsen & Soelensminde, 1998). However, it is not clear from this evidence alone which question format yields the more valid results.

One exploration of this issue was reported by Kealy and Turner (1993), who found that among respondents asked both an open and a closed measure of willingness to pay, responses to the open question were changed by asking the closed question before it. And O'Conor, Johannesson, and Johansson (1999) and Ready, Buzby, and Hu (1996), and Green, Jacowitz, Kahneman, and McFadden (1998) showed that asking a closed question first changes answers to a subsequent open question via anchoring: pulling responses to the open question toward those answers given to the closed question. In contrast, responses to the closed question are the same regardless of whether the open question precedes it or not (Kealy & Turner, 1993). This

suggests that the open question is more vulnerable to systematic measurement error due to question order than the closed question form.

Data from a few experiments can be used to compare the correlational validity of open and closed questions. The logic of this approach is that if theory suggests that WTP responses should be correlated with a particular predictor, then the strength of the relation between the two variables can be treated as an indicator of the validity of the measures involved. When comparing open and closed measures of WTP, the measure yielding the stronger correlations would presumably be the more valid one.

Three studies offering relevant data found no consistent differences between the two question formats in this regard. For example, Loomis (1990) showed that willingness to pay for protecting a California lake was correlated with various predictor variables about as strongly with an open measure of willingness to pay as with a closed measure. Brown et al. (1996) found that willingness to pay for a public good was similarly related to three criterion variables, regardless of whether willingness to pay was measured by an open or a closed question. And Kealey and Turner (1993) found attitudinal predispositions to be more strongly correlated with open measures of willingness to pay in one study but more strongly correlated with closed measures in a second study.

However, four other studies found that closed questions yielded results of apparently superior correlational validity. Brown et al. (1996) found that willingness to pay was associated with respondent income, as would be expected, using the closed measure but not the open measure. Bohara, McKee, Berrens, Jenkins-Smith, Silva, and Brookshire (1998) found the same strengthening of the association between willingness to pay and income using closed measures, as well as strengthening of the relations between willingness to pay and various attitudinal predictors. Loomis, Lockwood, and DeLacy (1993) found stronger associations of attitudinal predispositions and demographics with willingness to pay when the latter was measured by closed questions than when it was measured by open questions. Likewise, Boyle, Johnson, McCollum, Desvousges, Dunford, and Hudson (1996) found that willingness to pay to prevent oil spills was correlated with attitudes and media exposure, as would be expected, when willingness to pay was measured with a closed question but not when it was measured with an open question.

A similar conclusion is supported by studies exploring responsiveness of WTP judgments to manipulations of the quantity of the good being valued. For example, Loomis, Lockwood, and DeLacy (1993) asked respondents about their willingness to pay for public goods of various sizes and found that open and closed questions revealed comparable sensitivity to the scope of the good being purchased. However, the predictors of willingness to pay changed with the scope of the good, but this was only apparent using the closed question and not the open one. Likewise, Desvousges, Johnson, Dunford, Boyle, Hudson, and Wilson (1993) found that people were willing to pay more money to prevent a greater quantity of oil spills when willingness to pay was measured by a closed question, whereas this greater willingness was not apparent when an open question was used.

Another set of studies, focused on the correspondence of survey reports with observations of the same phenomena gauged using other methods, also generally supported the validity of the closed questions. For example, Walsh, Johnson, and McKean (1992) assessed the validity of the mean dollar values of various recreation activities to a group of people as assessed by open and closed questions. To do so, Walsh et al. (1992) compared the resulting mean dollar values with values generated by the "travel cost method," which involves inferring dollar value from people's actual behavior in performing the activities. The dollar value means suggested by closed survey question responses were closer to the travel cost means than were the means suggested by the open survey questions, thereby suggesting greater validity in the former. Seller, Stoll, and Chavas (1985) found the same result using the same method. Only Frykblom's (1997) study yielded the opposite results: open question measures of people's willingness to pay for a book were closer to what they actually paid than were closed question measures.

Taken together, these studies generally support the notion that closed questions yield higher mean values of WTP than do open questions, and the former appear to be more valid than the latter. To contribute further to this literature, we conducted an experiment in which respondents were randomly assigned to be asked either an open-ended question or a close-ended question measuring willingness to pay. The questionnaire was designed to incorporate various measurements and manipulations to permit gauging the validity of the willingness to pay responses gathered by the two methods. We begin below by describing the procedures implemented in the data collection, and then we describe the results we obtained and outline their implications for best practices in survey measurement.

## **Overview of Our Investigation**

This research will focus particularly on three fundamental questions regarding closed- versus open-ended questions in CV surveys: (1) Under what circumstances does question form generate differences in responses? (2) If there is a difference, what determines the direction of the difference? Which is more valid?

Some hypotheses to be tested include the following:

- Other things equal, closed-ended questions are easier to answer. For factual items that you know very well (e.g. the number of times you were married, the number of children, your street address etc) these are so easy to recall that it makes no difference which format is used. For factual items that are difficult to know or remember (e.g. how many movies did you see last year, how tall is the Empire State Building etc), it does make a difference, with the closed-ended format being significantly easier.
- 2) If you don't know the answer and are seeking a short-cut, the two formats offer different ways out. With the open-ended format, the tendency is to guess a "round" number 1, 5, 10, 100, or whatever. This response may be more like an ordinal ranking than a cardinal number. Indeed, the statistical model of ordinal ranking with unequally spaced truncation points may provide a good fit to the open-ended responses for people who were having difficulty with the answer. With the closed-ended format, the easy out may be to say

"yes". Thus, a statistical model with a background propensity to say yes may fit closedended responses for those who were having difficulty with the answer.

3) The norms of conversation are different with open and closed ended questions. Generally, an open-ended question is a pure question. A closed-ended question may be an assertion as much as a conventional question, inviting the answer "yes" or "no" depending on context ("Are you coming?" invites "yes." "Is this wrong of me?" invites "no.") In this context, whether or not the closed-ended question invites a particular response depends on who is presumed to have superior information, the questioner or the respondent. If the respondent presumes that the questioner has superior information – e.g. Is the population of Afghanistan more than 23 million? – there is more likelihood that he will assume that the answer to the question must be yes. This is less likely to happen if the respondent clearly has the superior information – Did you go to the movies more than 23 times last year? – or if the questioner specifically disclaims superior information before posing the question.

We examined how CV question format and structure affected survey responses both conceptually and empirically. At a conceptual level, we approached the issue from two perspectives -- survey theoretic and econometric. The survey research perspective has been lacking from recent discussions in the economic literature about closed- versus open-ended questions. The implicit assumption in this economic literature is that observed effects associated with changes in question form are peculiar to CV and would not arise in other types of questions, such as factual questions about purchasing behavior. The evidence, in fact, does not support this. There is, indeed, an immense literature in survey research on how question form affects survey responses dealing with open- versus closed-ended questions, as well as various alternative forms of closed-ended question. As part of this project, Krosnick analyzed the implications of survey research findings for the design and interpretation of CV experiments.

To the extent that there are differential response affects associated with open- and closed-ended questions, it may be possible to model these econometrically in a manner that captures the cognitive processes involved. Krosnick (1991) proposed a particular model of respondents to surveys as "cognitive misers" adopting various possible strategies that might make it easier for them to get through their participation in the survey. There is mounting evidence from recent work on cognitive survey design to support this hypothesis. However, the implications for statistical modeling of survey responses have barely been investigated by either survey researchers or econometricians. As part of this project, Hanemann has analyzed several possible econometric realizations of cognitive models of survey respondent behavior.

### RESULTS

## **Reanalysis of Prior Survey Studies**

In the next phase of our research, we focused on comparisons of open-ended with closed-ended questions to measure willingness to pay in surveys. In particular, we explored two charges that have been made against closed-ended questions: (1) they over-estimate willingness to pay, and (2) they do so because high bid amounts offered to respondents anchor their judgments at large values. To assess whether the latter assertion is true, we realized that we could conduct a number of informative analyses using existing large survey data sets that had been the basis of previous publications by various investigators. We therefore obtained these data sets, conducted analyses of them early in 1996 -1997, and stumbled onto a remarkable discovery.

We found that the results of previous investigations appear to hinge completely on a procedural detail that has thus far received little attention in this literature. Specifically, we realized that the results of an analysis using closed-ended willingness-to-pay questions hinges crucially upon the calculation method used to assess mean or median willingness-to-pay if the bid amounts offered to respondents are spaced far apart (as has been done in most past studies). When bid amounts are closely spaced, the two primary calculation methods yield very similar results. But when the bid amounts are spaced far apart, one calculation method systematically yields larger means and medians, and it is this method that has been used to date to support assertions that close-ended WTP questions yield overestimates.

We then reanalyzed previously-collected data using the alternative, and more defensible, approach. Our first reanalysis was of data reported by Green, Jacowitz, Kahneman, and McFadden (1995). These authors showed originally that closed-ended questions led to over-estimates of willingness to pay. But as Table 1 shows, when we reanalyzed their data, we found this not to be the case. They asked their respondents five questions, about how much money they would be willing to pay to prevent oil spills from harming seabids, how much money they would be willing to pay to reduce traffic accidents, as well as three questions not about willingness to pay (the height of the tallest redwood tree, average gasoline usage by car owners, and the amount of rainfall at Mount Wialeale).

The first column of numbers in Table 1 are the mean answers provided by Green *et al.*'s respondents when given an open-ended question. Under each mean is its standard deviation, and in the next column appears the range of values from one standard deviation below the mean to one standard deviation above the mean. Beginning in column 3 are estimates of the mean generated by different combinations of closed-ended questions. As is apparent there, no matter which set of anchors we used, the estimated mean response was rarely above the open-ended mean and was often below it. And for the one item where the mean estimated by the close-ended questions is significantly above that for the open-ended question (i.e., the rainfall question), the closed-ended results are closer to the true value of 451 than is the open-ended response.

Closed-Ended Questions											
Торіс	Open- Ended Mean	Open- Ended Range	Low 25,50	High 50,75	Leaning Low 25,50,95	Leaning High 25,90,95	Centered 25,50,75				
Seabirds	64 (13)	51-77	17.6	31.9	85.1	77.0	33.0				
Accidents	98 (22)	76-120	16.7	26.2	65.4	77.0	26.8				
Redwoods (true=366)	491 (108)	383-599	245.8	299.8	540.2	579.0	316.3				
Gasoline (true=56)	63 (7)	56-70	37.2	45.9	59.0	58.2	46.7				
Rainfall (true=451)	188 (21)	167-209	80.2	145.9	252.4	259.7	157.2				

#### Table 1: Reanalysis of Green et al. Data

Tables 2 and 3 present results estimated through reanalysis of data from the California Oil Spill survey study and the "Montrose" survey study on DDT and PCB contamination, both done by us in collaboration with our colleagues in the Natural Resource Damage Assessment group. We simply reanalyzed those data using only subsets of the closed-ended questions, and our interest again was in whether offering respondents only large dollar values can lead to estimates of mean willingness to pay that exceed results when more balanced sets of anchors are used. The first column of numbers in each table is the estimated mean WTP when all anchors are used, and the remaining columns display estimated means when only subsets of anchors are used. And never does the estimated mean using a subset exceed the result using all dollar values. If the alternative calculation method we used is indeed most reasonable, then these results suggest that anchoring does not appear to distort WTP estimates in a non-conservative direction.

Closed-Ended Questions											
	All Anchors	Very Low 10,25	Very High 140,215	Leaning Low 10,25,215	Leaning High 10,140,215	Centered 10,80,215					
Mean WTP	\$63	\$13	\$54	\$55	\$57	\$54					

#### Table 2: Reanalysis of Montrose Data

#### Table 3: Reanalysis of California Oil Spill Data

Closed-Ended Questions											
	All Anchors	Very Low 10,25	Very High 140,215	Leaning Low 10,25,215	Leaning High 10,140,215						
Mean WTP	\$76	\$14	\$68	\$63	\$69						

### **Experiments with Student Participants**

We have generated one additional set of empirical findings because of the question wording used by Green *et al.* to assess answers in their study of anchoring. These investigators asked respondents yes/no questions only emphasizing one side of the judgment to be made (e.g., "Would you be willing to pay \$25 to ..."), rather than asking balanced questions, emphasizing both sides equally (e.g., "Would you be willing to pay \$25 to ... or would you not be willing to pay \$25 for that purpose?"). This sort of wording is susceptible to acquiescence response bias, a tendency to say "yes" in answer to any question, regardless of its content.

If some respondents were indeed susceptible to this bias, it would have led the closed-ended questions used by Green *et al.* to yield over-estimates of WTP relative to their open-ended questions, just as they found. We therefore conducted two experiments in which we systematically varied the wording of the WTP questions across respondents.

The first set of data was collected from 445 undergraduate students enrolled in introductory psychology at Ohio State University during Spring quarter, 1996. Following the methodology of Jacowitz and Kahneman (1985) and Green, Jacowitz, Kahneman and McFadden (1995), participants in the calibration condition (N=51) estimated 16 unknown quantities. From the distribution of responses from these participants, values corresponding to the 15th and 85 percentile were selected for each quantity. These values were used an "anchors" in the experimental conditions. Participants in the experimental conditions (N=384) provided estimates

of the same 16 quantities. Before estimating each quantity, however, they made a judgment about an anchor value associated with each quantity.

The wording of these initial judgment questions was manipulated, and participants were assigned to one of four question wording conditions. Depending on which condition they had been assigned to, participants were asked:

"Is (<u>quantity X</u>) greater than (<u>anchor value</u>)?"

"Is (quantity X) less than (anchor value)?"

- "Is (quantity X) greater than or less than (anchor value)?"
- "Is (quantity X) less than or greater than (anchor value)?"

The anchor values were also manipulated: participants were assigned to either the low anchor condition (all initial anchor values corresponded to the 15th percentile of the calibration group's estimates) or the high anchor condition (all initial anchor values corresponded to the 95th percentile of the calibration group's estimates).

After they made this initial judgment for a quantity, they then provided their own estimate of the quantity. Finally, participants rated their level of confidence in their estimate on a 10-point scale.

#### Study 2

The second set of data was collected from 667 undergraduate students recruited at the Ohio Union at the Ohio State University and paid \$3.00 to participate. The methodology was identical to that reported above, with one exception: In this study, all participants were presented with low anchors for half of the items and high anchors for the other half.

These two studies' datasets were analyzed to explore the hypothesis that the results obtained from closed-ended survey questions may be upwardly biased due to "yea-saying," in which a percentage of respondents answer "yes" irregardless of the bid values. Hanemman hypothesized that the degree of "yea-saying" may be correlated with a respondent's uncertainty; the greater the uncertainty, the larger the probability of "yea-saying."

In order to test this hypothesis, Krosnick and Hanemann analyzed the data from the OSU experiments. Respondents were asked to state "yes" or "no" whether the true value of an item was greater than or less than a stated bid value. Three types of questions were asked: 1) questions for which the respondent would likely have little prior knowledge or information (i.e. the speed of a house cat or the average amount of rainfall in Hawaii), 2) questions for which the respondent would likely have some knowledge (i.e. how many students are at OSU or how many bars are in Ohio), and 3) questions for which the respondent would have perfect or near perfect information (i.e. how many brothers and sisters they have or how many miles they have driven in the last 14 days). Each respondent was asked the same 24 questions. An equal number of respondents were presented with each question in a less than format (i.e. "Is the true value of X less than Y?") and a greater than format, using six different percentile bids for Y, generated from an earlier sample of open-ended responses to the same questions.

The data from this experiment provides a formal test of yea-saying, because in the absence of "yea-saying," at each percentile bid, the sum of the percent of "yes" responses for the less than format plus the percent of "yes" responses from the greater than format should equal 1. If they sum to significantly greater than 1, that would be evidence of yea-saying, since it indicates a significant number of "yes" responses that are not sensitive to the bid value. Hanemann's hypothesized that "yea-saying" should not occur for question type 3 (where the respondent is highly certain) but may occur in question type 1 (where the respondent is highly uncertain).

In fact, the study did not produce evidence of "yea-saying." If anything, there is a preponderance of "no" responses in many of the questions (i.e. "nay-saying"). Table 1 shows the mean "yes" response rate at each percentile bid for the three types of questions in the OSU survey, for both the greater than format and the less than format. Table 1a contains the results of the t-tests of whether the sum of the "yes" responses across question wording formats equals 1 at each percentile bid. For 5 bids, the sum of the "yes" responses is significantly less than 1, while the rest are not statistically different than 1.

Greater Than Format								Les	s Than	Forma	<u>t</u>		
Percent							centil	es					
Question	$8^{\text{th}}$	$18^{\text{th}}$	38 <sup>th</sup>	62 <sup>nd</sup>	$82^{nd}$	92 <sup>nd</sup>		$8^{\text{th}}$	$18^{\text{th}}$	$38^{\text{th}}$	$62^{nd}$	$82^{nd}$	92 <sup>nd</sup>
Type													
1	.82	.78	.66	.56	.46	.36		.15	.21	.31	.37	.49	.60
2	.84	.78	.76	.55	.37	.29		.16	.21	.25	.46	.58	.72
3	.71	.60	.47	.58	.43	.16		.29	.44	.50	.53	.63	.77

 Table 4: Mean "Yes" Responses at Each Percentile Bid by Question Type

 Table 4a: "Yes" Responses From Greater Than Format + "Yes" Responses from Less than Format =1

	Percentiles											
Question	G8+L8=1	G18+L18=1	G38+L38=1	G62+L62=1	G82+182=1	G92+L92=1						
Type												
1	Can't	<1*	<1**	Can't	<1**	Can't						
	Reject			Reject		Reject						
2	Can't	Can't	Can't	Can't	Can't	Can't						
	Reject	Reject	Reject	Reject	Reject	Reject						
3	Can't	Can't	<1**	Can't	Can't	<1**						
	Reject	Reject		Reject	Reject							

\*=90% confidence,\*\*=95% confidence

Although there is no evidence of "yea-saying" in the OSU data, there are some other interesting patterns. Table 2 shows the difference between the percentage of "yes" responses at each percentile bid (by question type) and the "true" percentile (garnered from the sample of open-ended responses). For example, for question type 1 at the bid representing the 8<sup>th</sup> percentile, 92% of the respondents "should have" indicated that the value of the item was greater than the bid. In fact, only 82% answered "yes," so the divergence from the true probability is -.10.

Greater Than Format									Less 7	Гhan F	ormat			
Percentiles														
Question	8 <sup>th</sup>	$18^{\text{th}}$	38 <sup>th</sup>	$62^{nd}$	$82^{nd}$	92 <sup>nd</sup>	Mea	$8^{\text{th}}$	$18^{\text{th}}$	$38^{\text{th}}$	$62^{nd}$	$82^{nd}$	92 <sup>nd</sup>	Mean
Туре							n							
1	10	04	.04	.18	.28	.28	.11	.07	.03	07	25	33	32	14
2	08	04	.14	.17	.19	.21	.10	.08	.03	13	16	24	20	10
3	21	22	15	.00	.25	.08	04	.21	.26	.12	09	19	15	03
Mean	13	10	.01	.12	.24	.19		.12	.11	03	17	25	22	

 Table 5: Difference Between Percentage of "Yes" Responses And The True "Percentiles"

In the "greater than" format for question type 1, respondents answered "no" slightly more than expected for lower valued bids, but when the bids increased they answered "yes" much more than expected. For the "greater than" format in question type 3, the results are different; respondents answered "no" much more than expected at the lower bids, but at the higher bids they didn't exhibit such a strong preponderance of "yes" responses. In the "less than" format, these results are reversed; in question type 1, there is a slight preponderance of "yes" responses at the lower bids and a high preponderance of "no" responses at the highest bids, but not as many "no" responses at the higher bids.

The means of the differences, both at each percentile and each question type (across all percentiles) exhibit interesting patterns as well. Overall, in the "greater than" format there is an excess of "no" responses at lower bid values and an excess of "yes" responses at high bid values. This pattern is reversed in the "less than" format.

The mean of the differences across bids is greatest for question type 1 and least for question 3 (in absolute value) in both question wording formats; .11 vs. .04 in the greater than format and .14 vs. .03 in the less than format respectively. Although these divergences largely cancel each other out when summing the percentages of "yes" responses across wording formats (producing the results in Table 1a), overall, responses to question type 3 (where the respondent has the least uncertainty) are the closest to the "true" values, while responses to question type 1 exhibit the greatest overall divergence from the "true" values.

Although the hypothesis that the OSU data would show a greater degree of "yea-saying" in type 1 questions versus type 3 questions was not substantiated, there is evidence that different thought processes were at work, both across question types and question wording formats. At the present,

the authors are investigating what psychological phenomena might be able to best explain the patterns in the OSU data and how to model them.

#### Simulation

In order for us to proceed with our investigation, we must understand the properties of the two estimators and their impact on comparisons of open-ended and closed-ended questions. We suspect that: (1) given a particular set of design points, one estimator may indeed be statistically superior to the other in terms of precision and robustness of estimation; and (2) given the welfare measure one wishes to calculate, the design points for a CV survey must be chosen with care in order to maximize the efficiency of the estimate resulting when the data have been collected. There is a small literature on item (2), which starts with Barbara Kanninen's Ph.D. dissertation written under Michael Hanemann's supervision here in 1991. But there has been almost no investigation of (1). To remedy this, we decided that we must pause in our investigation and conduct a Monte Carlo simulation analysis demonstrating the impact of various design decisions and data structures on estimates of WTP.

Only when it is completed can we have the confidence that we have properly understood the dynamics we have observed in our analyses to date, and only then can we confidently finalize the design of our final survey experiment. This is so because if we are correct that results hinge both on calculation method and on bid spacing, then we will incorporate experimental manipulations of bid spacing in our final study. This aspect of the design was not in our original plans but now may be essential.

### National Survey

To design our national survey, we conducted extensive research to locate an environmental problem in Ohio suitable for our study. After completing such archival research, we concluded that a suitable instance would be PCB contamination of Ohio rivers. We therefore drafted a survey question that would describe this situation, propose a procedure for repairing the problem, and measuring respondent willingness to pay for a such a cleanup effort.

This question was asked in a telephone survey of a representative sample of adult residents of the State of Ohio during the late Fall of 1998 and the Winter of 1999.

The Center for Survey Research at Ohio State University collected data for a telephone survey experiment comparing two alternative contingent valuation survey formats based on open-ended versus closed-ended elicitation of willingness to pay (WTP). A total of approximately 3,000 adult residents of Ohio were interviewed. During the course of the interviews, which lasted 6 minutes on average, respondents were told about pollution of rivers in Ohio and were told about a proposed plan to remedy this. They were then asked their willingness to pay to implement the plan, using one of two alternative survey formats. Some respondents were asked an open-ended question requesting them to say what is the most they would be willing to pay in a one-time extra state tax payment to help cover the costs of implementing this plan. Other respondents were

instead told the cost to their household of implementing the plan, in the form of a one-time extra state tax payment, and these respondents were asked whether they would vote for or against the plan if given an opportunity to do so. Among the respondents asked this closed-ended question, different ones were randomly assigned to be told different costs ("bid amounts") to their households.

The purpose of the survey was both to compare the results obtained with the two formats and also to test several hypotheses about factors that might be responsible for differences in the two formats' performance.

Two main hypotheses about the open-ended format are that respondents find it more difficult to answer than the closed-ended format and that their responses to the open-ended format are influenced in a downward direction by their assessment of what the program might cost. To test the first hypothesis, the survey allowed for a response of "don't know" in both the open- and closed-ended formats. To test the second hypothesis, after the valuation question respondents where asked for their best guess of how much it would cost to implement the clean-up plan.

An issue related to the closed-ended format is whether or not respondents engage is the effect of response choice order. As a result of satisficing in dealing with a survey interview, some respondents choose the last option that is read to them during telephone interviews. To test whether this occurs during CV interviews with closed-ended WTP measures, we built in another experimental manipulation, For each bid amount, a split-sample comparison was conducted using two variants of the closed-ended elicitation question: one variant asked respondents whether they would vote for or against the program, while the other reversed the sequence and asked whether they would vote against or for the program. For respondents who are simply choosing the last option they hear, reversing the response sequence should induce a reversal of WTP responses.

Another issue with the closed-ended format is the sensitivity to the dollar amounts (bids) employed. To investigate this, the closed-ended survey employed 16 distinct bids, ranging from \$15 to \$240 per household. By sub-sampling the data, it was possible to ascertain how the results would change if only subsets of these bids had been employed.

Crossed with these manipulations of question format and structure was a manipulation of scope. Some respondents (randomly selected) were told that the pollution is contaminating three Ohio rivers, whereas other respondents were told it is contaminating only one river in Ohio. The hypothesis tested is that there would be a larger WTP to eliminate pollution in three rivers than in one.

All respondents were asked a series of questions that we expected would be predictors of their willingness to pay, including their perceptions of the seriousness of the damage being caused by the pollution, their perceptions of the repair plan's likely effectiveness, their general support for environmental protection, their household incomes, and so on.

When respondents were asked whether they would vote for or against a one-time tax increase to clean up PCB deposits in Ohio rivers, many Ohio residents said they would vote in favor of such

an increase. Likewise, when other respondents were asked how much they would be willing to pay as a one-time tax increase for clean-up efforts, many people said that they would be willing to pay at least something. However, there was a sizable difference in the amount that respondents said they would be willing to pay between these two question forms. The mean WTP estimate from closed-ended question responses was over three times larger than the average amount respondents said they would be willing to pay in response to an open-ended question. Which of these WTP estimates is likely to be the most valid? It is difficult to answer that question from these data alone, because responses to the two question forms were both predicted by the full array of theoretically sensible variables. As long as the sample size was sufficiently large to detect effects, and as long as steps were taken to account for skewness in responses to the open-ended question during statistical analysis, WTP amounts were sensitive to the scope of the problem, perceived effectiveness of the plan to remedy the problem, environmentalism attitudes, and income as would be expected if respondents were thinking sensibly about the value of the good.

### CONCLUSION

These results lend support to the argument that CV questions are valid measures of the public's willingness to pay for environmental goods, despite some critics' objections that such measures are seriously lacking in validity. Respondents were sensitive to key components of CV questions (e.g., scope of the problem), and also sensibly based their judgments on their personal beliefs and their ability to surrender money.

However, it is not clear why mean WTP estimates differed so much, since responses to both types of questions appear to manifest correlational validity. This mystery remains to be investigated further. If people responding to closed-ended questions anchored on the bid amount mentioned by the question, they appear not to have done so mindlessly, since these values can be significantly predicted by other variables expected to capture true underlying preferences.

The present research adds to the existing literature on open and closed CV question specifically, and open vs. closed survey questions for measuring attitudes more generally. Though some past work on the difference between types of WTP questions has come to the conclusion that closed-ended questions yield the highest quality data, the present study suggests that both open and closed questions can be valid. It seems necessary to review this past work to see if there are consistent practices in design and analysis that might favor one question form. For example, did the same size samples provide responses to both question forms, and were attempts made to account for skewness in the open-ended data? The present study suggests that close attention should be paid to such factors in order to have a clearer picture of the validities of these two survey methods.

Though in this study both open and closed questions seemed to be equally valid, researchers should be cautioned not to uncritically generalize these results to all open and closed survey questions. There are many reasons to think that CV questions, regardless of whether they are open or closed, are quite different than other attitude questions. Open CV questions clearly imply a metric that respondents understand well and are quite familiar with (e.g., dollars). Other open questions that have been studied in the past (e.g., "What is the most important problem facing the country today?") typically allow respondents to provide a seemingly endless array of possible answers. Also, the referendum format of closed-ended CV questions stands in contrast to other closed-ended questions, because CV survey respondents simply say "yes" or "no" with regard to a particular proposition. Non-CV closed-ended questions, can take on many more forms than do CV questions.

Although it might not be appropriate to broadly generalize these results to all survey questions, this caution is not meant to undermine the contribution of this work. Although CV questions are unusual, they are often used for gathering information about the economic value of public goods to individual citizens. Our work therefore contributes to a general understanding and mastery of best practice in the design and implementation of contingent valuation surveys, and we hope will spark further interest in how to design optimal questionnaires for measuring attitudes and beliefs.

#### REFERENCES

- Alwin, D.F. and J.A. Krosnick. 1985. The Measurement of Values in Surveys: A Comparison of Ratings and Rankings. *Public Opinion Quarterly* **49**: 535-552.
- Arrow, K., R. Solow, P. Portney, E. Learner, R. Radner, and H. Schuman. *Report of the NOAA Panel on Contingent Valuation.*, Washington, D.C.: Jan., 1993.
- Bishop. R.C., and T.A. Heberlein. 1979. Measuring Values of Extramarket Goods: Are Indirect Measures Biased? *American Journal of Agricultural Economics*, **61**: 926-930.
- Ciriacy-Wantrup, S.V. 1947. Capital Returns from Soil-Conservation Practices. *Journal of Farm Economics*, **29:** 1188-90.
- Cummings, R.G., D.S. Brookshire, and W.D. Schulze. *Valuing Environmental Goods: An* Assessment of the Contingent Valuation Method, Totowa, New Jersey: Rowman and Allanheld, 1986.
- Davis, R.K. 1963. *The Value of Outdoor Recreation: An Economic Study of the Maine Woods.* Dissertation. Harvard University.
- Hanemann, W.M. 1992. Natural Resource Damages for Oil Spills in California. In Kevin M. Ward and John W. Duffield, *Natural Resource Damages: Law and Economics*. New York: John Wiley & Sons, Inc., 555-580.
- Hanemann, W.M. 1984. Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses. *American Journal of Agricultural Economics*, August, **66:** 332-341.
- Hanemann, W.M., J. Loomis, and B. Kanninen. 1991. Statistical Efficiency of Double-Bonded Dichotomous Choice Contingent Valuation. *American Journal of Agricultural Economics*, November, **73**: 1255-1263.
- Hanemann, W.M., and J. Cooper. Referendum Contingent Valuation: How Many Bounds Are Enough? report presented at the annual AAEA meeting, San Diego, CA, August. 1994.
- Kanninen, B. 1993. Bias in Discrete Response Contingent Valuation. Washington, D.C: Resources for the Future. Discussion Paper QE93-27, September.
- Krosnick, J.A. 1991. Response Strategies for Coping with Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, **5**: 213-236.
- Krosnick, J.A. Designing Questionnaires to Measure Attutudes. New York: Oxford University Press, In preparation.
- Loomis, J.B., W.M. Hanernann, B. Kanninen, and T. Wegge. 1991. Willingness to Pay to Protect Wetlands and Reduce Wildlife Contamination from Agricultural Drainage. In *The Economics and Management of Water and Drainage in Agriculture* edited by Ariel Dinar and David Zilberman, Boston: Kluwer Academic Publishers, 411-429
- Mitchell, R.C., and R.T. Carson. Using Surveys to Value Public Goods: The Contingent Valuation Method, Washington, D.C.: Resources for the Future, 1989.

Randall, A., B.C. Ives, and C. Eastman. 1974. Bidding Games for Valuation of Aesthetic Environmental Improvements. *Journal of Environmental Economics and Management*, 1: 132-149.



#### The Department of the Interior Mission

As the Nation's principal conservation agency, the Department of the Interior has responsibility for most of our nationally owned public lands and natural resources. This includes fostering sound use of our land and water resources; protecting our fish, wildlife, and biological diversity; preserving the environmental and cultural values of our national parks and historical places; and providing for the enjoyment of life through outdoor recreation. The Department assesses our energy and mineral resources and works to ensure that their development is in the best interests of all our people by encouraging stewardship and citizen participation in their care. The Department also has a major responsibility for American Indian reservation communities and for people who live in island territories under U.S. administration.



#### The Minerals Management Service Mission

As a bureau of the Department of the Interior, the Minerals Management Service's (MMS) primary responsibilities are to manage the mineral resources located on the Nation's Outer Continental Shelf (OCS), collect revenue from the Federal OCS and onshore Federal and Indian lands, and distribute those revenues.

Moreover, in working to meet its responsibilities, the **Offshore Minerals Management Program** administers the OCS competitive leasing program and oversees the safe and environmentally sound exploration and production of our Nation's offshore natural gas, oil and other mineral resources. The MMS **Royalty Management Program** meets its responsibilities by ensuring the efficient, timely and accurate collection and disbursement of revenue from mineral leasing and production due to Indian tribes and allottees, States and the U.S. Treasury.

The MMS strives to fulfill its responsibilities through the general guiding principles of: (1) being responsive to the public's concerns and interests by maintaining a dialogue with all potentially affected parties and (2) carrying out its programs with an emphasis on working to enhance the quality of life for all Americans by lending MMS assistance and expertise to economic development and environmental protection.