RTI Press

Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy

William H. Desvousges, F. Reed Johnson, Richard W. Dunford, Kevin J. Boyle, Sara P. Hudson, and K. Nicole Wilson

Second Edition

RTI INTERNATIONAL

Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy

William H. Desvousges, F. Reed Johnson, Richard W. Dunford,
Kevin J. Boyle, Sara P. Hudson, and K. Nicole Wilson

Second Edition

RTI Press

RTI INTERNATIONAL

RTI Press publication BK-0001-1009

Originally published in 1992 as Research Triangle Institute Monograph 92-1. This second edition differs from the original in the following ways: it has a new publisher and an additional Foreword and Preface. The original text is unchanged, and no additional analyses were performed on the data.

Library of Congress Control Number: 2010935161
ISBN: 978-1-934831-00-7

https://doi.org/10.3768/rtipress.2009.bk.0001.1009
www.rti.org/rtipress

About the Authors

At the time of original printing, **William H. Desvousges**, PhD, was an economist in Research Triangle Institute's (RTI's) Natural Resource Damage Assessment Program in the Center for Economics Research.* In 2005 he formed W. H. Desvousges & Associates, Inc., an environmental resources consulting firm, and he remains its president.

At the time of original printing, **F. Reed Johnson**, PhD, was an economist in RTI's Natural Resource Damage Assessment Program. He is now a Senior Fellow and principal economist in RTI Health Solutions, a business unit of RTI.

Richard W. Dunford, PhD, was also an economist in RTI's Natural Resource Damage Assessment Program. In July 2005, he founded Environmental Economics Services, LLC, an independent economics consulting firm, to assist businesses in applying economics to complex environmental and natural resource issues.

Kevin J. Boyle, PhD, worked for the University of Maine Department of Agricultural and Resource Economics in 1992, and he has continued in the field of economic research. He is currently professor and head of the Department of Agricultural and Applied Economics at Virginia Polytechnic Institute and State University (Virginia Tech).

At the time of original printing, **Sara P. Hudson** was an economist in RTI's Natural Resource Damage Assessment Program. She is now a school teacher.

Also an economist in RTI's Natural Resource Damage Assessment Program, **K. Nicole Wilson** is now the Ronald McDonald House coordinator for Southern California.

* Research Triangle Institute is now doing business as RTI International.

This publication is part of the RTI Press Book series.

RTI International
3040 Cornwallis Road, PO Box 12194, Research Triangle Park, NC 27709-2194 USA
rtipress@rti.org
www.rti.org

Contents

Appendices

Tables

List of Figures

Foreword

Kevin J. Boyle, Professor and Head
Department of Agricultural and Applied Economics
Virginia Tech

When the *Exxon Valdez* oil spill occurred on March 24, 1989, the application of contingent valuation to estimate use values, let alone nonuse values, was still in the early stages of development. Only two nationally prominent contingent-valuation studies had been conducted to estimate values for changes in water quality (Mitchell and Carson, 1981; Desvousges, Smith, and McGivney, 1983), and neither study valued the consequences of a major oil spill in a marine environment. Mitchell and Carson (1993) valued the national benefits of freshwater pollution control. Desvousges and colleagues estimated values for enhanced water quality in the Pennsylvania portion of the Monongahela River (Desvousges, Smith, and Fisher, 1987).

Three years prior to the spill, Cummings, Brookshire, and Schulze (1986) had published a state-of-the-art assessment of the contingent-valuation method where they suggested the reference accuracy of contingent-valuation estimates might be ±50% of the true value. The problem is that the true value can never be observed, and parallel experiments are required to infer accuracy. Two studies at that time had investigated the comparability of contingent-valuation estimates to estimates derived using cash transactions. These are tests of criterion validity where the cash-transaction estimates are the criterion, presumed truth, by which the contingent-valuation estimates are judged. Heberlein and Bishop (1986) found biases of 25, 33, and 0 percent in three experiments selling a good not typically sold in markets, deer hunting permits for a hunting area in central Wisconsin. Dickie, Fisher, and Gerking (1987), in a field experiment selling a market good (strawberries), found no significant difference between the contingent-valuation estimate and the result

Note: In this foreword, I have added citations from the literature as examples but do not claim to represent an exhaustive review of the literature. The spirit of this foreword is to place the studies presented in this monograph in the context of the stated-preference valuation literature in 1990 and 2010.

from cash transactions. These studies were available to Cummings, Brookshire, and Schulze and suggest two possible inferences that might modify the ±50% accuracy: (1) the error might be lower than |50%|, and (2) when errors occur they are likely to be overestimates (>0).

There is more to consider, however, that might lead one to question whether overestimation errors in contingent-valuation studies might exceed Heberlein and Bishop's 25 and 33 percent errors. Both validity studies were conducted in contexts where one would expect contingent valuation to work "best"; both studies estimated use. The hunters in the Heberlein and Bishop study likely had considerable prior experience making choices of when and where to hunt that involved differing levels of out-of-pocket expenditures. Strawberries are a market good that most people are likely to have transaction experience with.

At the time of the spill, this foundation for investigating the validity of contingent-valuation applications to use values presented challenges when valuing changes in water quality where nonuse values are involved and people do not have direct or indirect transaction experience. For example, what was the accuracy of the Mitchell and Carson national freshwater quality study? Purchasing improvements in water quality is not something that people have experience with, and this study elicited what has become known as a total value, which includes use and nonuse values. The criterion validity studies cited above cannot speak to the accuracy of the Mitchell and Carson study.

Give this status of the academic literature and the expectation that the resource trustees for the damages from the *Exxon Valdez* oil spill would develop a total value monetary estimate of resource injuries based on a contingent-valuation study, Exxon hired the research group at Research Triangle Institute, of which I was a member, to investigate issues related to the use of contingent valuation. The research inquiries arose from two key questions. Are contingent-valuation estimates of nonmarket values valid? Are contingent-valuation estimates of injuries to aquatic ecosystems from oil valid? There was never any intent to value the injuries from the *Exxon Valdez* oil spill. Rather, the intent was to investigate issues that were fundamental to contingent valuation and might arise in any contingent-valuation estimate of monetary injuries developed by the resource trustees.

As with any area of research inquiry, there are numerous lines of investigation that can be undertaken. The Research Triangle Institute team, without any pressure or guidance from Exxon, chose three lines of investigation.

The first line of inquiry was empirical investigations of theoretical validity that considered the sensitivity of total value estimates to the extent of injury. One might logically expect that larger oil spills would result in larger resource injuries. Two experiments were undertaken.

- One study investigated how total values would differ between a program that reduced the effects of small spills (<50,000 gallons) versus a program that reduced the effects of all spills (small and large spills). These small and large spills covered actual spill amounts that had occurred. The *Nestucca* barge spill released about 230,000 gallons of fuel oil.[1] The Arthur Kill pipeline leak released about 570,000 gallons of fuel oil.[2] The *Exxon Valdez* spill released about 11 million gallons of crude oil.[3] The idea for this study arose from the Oil Pollution Act of 1990.[4]
- The second study investigated how total values would differ with three levels of bird deaths (2,000, 20,000, and 200,000). These numbers corresponded to the range of actual bird deaths that had been observed from oil spills. For example, the New Jersey Audubon Society reported that the Arthur Kill pipeline leak had "dire impact on hundreds of birds."[5] The State of Washington reported that the *Nestucca* spill killed or injured 56,000 seabirds.[1] Sarah Graham reported in *Scientific American* that as many as 250,000 bird deaths were attributable to the *Exxon Valdez* oil spill.[6]

---

[1] http://www.ecy.wa.gov/programs/spills/incidents/Nestucca/NestuccaHistory.pdf, accessed June 6, 2010.

[2] http://www.darrp.noaa.gov/pdf/HEP_Article_Autumn08.pdf, accessed June 6, 2010.

[3] http://www.evostc.state.ak.us/facts/qanda.cfm, accessed June 6, 2010.

[4] http://www.epa.gov/oem/content/lawsregs/opaover.htm, accessed June 8, 2010.

[5] http://www.njaudubon.org/SectionConservation/NJASOpinion/OilSpillontheArthurKill.aspx, accessed June 6, 2010.

[6] Graham, Sarah (December 19, 2003). "Environmental Effects of *Exxon Valdez* Spill Still Being Felt." *Scientific American*. http://www.sciam.com/article.cfm?chanID=sa003&articleID=0001A1FF-12D7-1FE2-92D783414B7F0000http://en.wikipedia.org/wiki/Exxon_Valdez_oil_spill. Retrieved March 9, 2008, accessed May 28, 2010.

If injuries are linked to the size of an oil spill and bird deaths are one of the primary injuries, it seemed logical to investigate whether the magnitude of a spill and the magnitude a key component of injury influenced the magnitude of value estimates. These types of investigations have become known as tests of scope.

At the time of the spill, Kahneman (1986) had put forth the argument, with some weak empirical evidence, that contingent-valuation estimates might not be sensitive to the extent of injury. The studies described above were designed to investigate the issue of sensitivity to scope more rigorously. These scope investigations of theoretical validity were based on the hypothesis:

$$H_0: \left|\frac{\partial U(\cdot)}{\partial Q}\right|_{\Delta Q_i} > \left|\frac{\partial U(\cdot)}{\partial Q}\right|_{\Delta Q_j}, \Delta Q_i > \Delta Q_j \rightarrow wtp_i > wtp_j$$

$$\text{vs.}$$

$$H_A: \text{not } H_0,$$

where Q could be quantity of oil spilled or quantity of birds killed, and wtp is a total value measure of willingness to pay for the change. This simple hypothesis followed the theoretical presumption that larger spills or larger injuries will result in larger decreases in utility and, therefore, larger willingness to pay to avoid injury.

The second line of inquiry asked if different formats of the contingent-valuation question result in statistically similar estimates of total value. The oil-spill sample, noted above, was randomly stratified into two subsamples for both the small spills and all spills treatments. One subsample from each spill-size treatment was administered an open-ended question and the other was administered a dichotomous-choice question. The hypothesis was:

$$H_0: wtp_{oe} = wtp_{dc}$$

$$\text{vs.}$$

$$H_A: \text{not } H_0,$$

where oe denotes an estimate from an open-ended question and dc denotes an estimate from a dichotomous-choice question. This is a test of convergent validity; two procedures measuring the same theoretical construct should provide statistically similar estimates of value.

At the time of the oil spill there were four common variants of contingent-valuation questions (open-ended, iterative bidding, dichotomous choice, and payment cards), and two major studies had compared selected types of these formats. Desvousges, Smith, and Fisher (1987), using data from the Monongahela study cited above, compared option price estimates of use values for iterative bidding, open-ended, and payment card question formats. They concluded that iterative bidding "with a $125 starting point and the payment card approach appear to have led to higher responses than the other two formats" (p. 265). Boyle and Bishop (1988) found iterative bidding and payment cards provided statistically similar estimates of the effect of a change in river scenery quality on use values, and that a dichotomous-choice question provided a statistically smaller estimate of value. These studies suggest that contingent valuation question formats do not necessary satisfy tests of convergent validity.

Iterative-bidding questions were becoming discredited due to starting-point bias (Boyle, Bishop, and Welsh, 1985; Samples, 1985; Thayer, 1981). Payment cards were guilty by association with the iterative-bidding framework where questions were raised, but not answered, about how bid amounts on payments cards might affect estimates of willingness to pay. Thus, payment cards never experienced broad acceptance in the literature. This left open-ended questions, used in the first known contingent valuation study (Davis, 1963), and dichotomous-choice questions, introduced in the Heberlein and Bishop (1986) study, as the two primary questions formats in the literature at the time and they were chosen for investigation.

The third and final investigation considered the reliability of dichotomous-choice estimates of total value, and the key element was whether respondents were responsive to high bid amounts. That is, the probability that a person would answer yes to a dichotomous-choice questions is a function of the bid amount posited in the question:

$$\Pr(\text{Yes}) = \Pr[U(Q_i,I) \geq U(Q_j, I\text{-}B)],\ \frac{\partial U(\cdot)}{\partial Q} < O \text{ and } Q_i > Q_j$$

where Q is as defined above, I is income and B is the bid amount in the dichotomous-choice question. Since wtp is a finite amount that is bounded by income, and the environmental quality being valued is not a necessary good, it is assumed that:

$$\text{b}\lim_{B \to I} \Pr(\text{Yes}) = 0.$$

A well-known problem among practitioners, but seldom documented, was the phenomenon known then as “fat tails”: the probability of a “yes” response stays very high and asymptotically approaches a non-trivial percentage as bid amounts increase. No statistical tests were done here, but the analysis asked how sensitive estimates of means and standard errors of wtp were to removing the largest bid ($1,000) from the econometric analyses of the small spill and all spills response data for the dichotomous-choice question format.

Each of these three investigations may seem overly simplistic in hindsight. The key issue in designing the studies is that they were being prepared in anticipation of litigation. While Exxon did not exert influence on the choice of study topics nor how the studies were conducted, we were acutely aware that the study findings might be used in court at some future date. It would be necessary to explain the issues tested in simple terms to a judge who had no knowledge of the nonmarket-valuation literature or to a jury who would have very limited knowledge of economics. Thus, simple questions were posed that addressed fundamental concerns in the contingent-valuation literature and would be intuitive to a lay person.

From an academic perspective, there were three journal articles published that were totally or in part based on the research reported in this monograph (Boyle et al., 1994; Boyle, Johnson and McCollum, 1997; Boyle et al., 1996). The results of all three of the investigations, discussed in detail in this monograph, suggested that problematic issues remained with contingent valuation and set off a variety of debates in the economics literature. I would argue that the results of these study investigations motivated:

- more careful design of stated preference scenarios to enhance the potential to show sensitivity to the scope of change being valued;
- the shift in the literature from contingent valuation to choice-modeling approaches to implementing stated-preference studies;
- more careful design of dichotomous-choice questions and the bid designs used in these questions; and
- alternative approaches to analyzing dichotomous-choice response data.

While the outcome of the research reported was not popular among practitioners and very popular among critics of contingent valuation, I think it is safe to say that this research had a lasting, positive impact on the application of contingent valuation and stated-preference methods in general.

Fast forward a little more than 21 years from the *Exxon Valdez* oil spill and the U.S. is facing an oil spill in the Gulf of Mexico that is much larger than the *Exxon Valdez* oil spill. There is no doubt that there will be a natural resource damage assessment, and contingent valuation or an alternative stated-preference method is again likely to play an important role in the monetary measurement of injuries. There has been considerable research on the validity of contingent valuation, and more generally stated-preference valuation methods, in the intervening years. Summary insights from this research address issues raised 20 years ago, when the research for this monograph was proposed and undertaken:

- More is known about the accuracy of contingent valuation now than 20 years ago. Murphy et al. (2005) reviewed 28 stated-preference studies that compared hypothetical transactions, such as contingent-valuation experiments and experimental auctions, with parallel cash transactions. They found a median overestimation bias of 35 percent, which is only slightly above the biases reported by Heberlein and Bishop (1986) nearly 25 years ago. In addition, researchers have demonstrated approaches such as income and substitute reminders that can be used to reduce overestimation bias in stated-preference studies (Kotchen and Reiling, 1999; Loomis, Gonzalez-Caban and Gregory, 1994), cheap talk (Cummings and Taylor, 1999) and uncertainty scales (Champ et al., 1997).
- Identifying whether estimated values are responsive to scope in contingent-valuation studies remains a difficult task (Heberlein et al., 2005; Veisten et al., 2004). A fundamental challenge in many contingent-valuation tests of scope is that between-subject, external tests of scope are conducted. This is the approach in the experiments reported in this monograph. In these experiments, where subjects see only one level of injury, they are asked to meet a level of rigor in their experimental choices that they are not asked to meet for purchase decisions they make in markets. In markets, people see the alternatives they are choosing among and can make relative decisions on what alternative is most desirable and the marginal value of the alternative they choose. Within the stated-preference

literature, there has been a trend away from contingent-valuation experiments to choice experiments (variously known as choice studies, choice modeling, attribute-based questions, or conjoint analysis) where respondents see the different levels of injury (or environmental quality) they can choose among (Holmes and Adamowicz, 2003). Choice experiments provide the basis for between-subjects, internal tests of scope that more closely mimic alternatives people face when making market decisions. Foster and Mourato (2004), for example, found that choice experiments exhibited more sensitivity to scope than was observed in a parallel contingent-valuation experiment.

- The NOAA Blue Ribbon Panel (Arrow et al., 1993) strongly recommended the use of referendum questions (dichotomous-choice questions framed as a referendum) for implementing contingent-valuation studies, and this has become the norm in the literature. The variants in question framing are now single, one-and-one-half, double, and multiple-bounded referendum questions (Alberini and Boyle, 2003; Bateman et al., 2009). Thus, the other formats for asking contingent-valuation questions are now rarely used.
- Referendum (dichotomous-choice) questions still face issues of fat tails, but nonparametric methods are now used to address this issue in data analyses (Haab and McConnell, 1997). The nonparametric approaches also reduce overestimation bias because the upper end of the response distribution is truncated in these analyses.

These summary points indicate that much has changed in 21 years, and much more is known about the accuracy of contingent valuation and stated-preference methods today than was the case in 1989, but there is still more to learn. A median overestimation bias in contingent-valuation studies satisfies the Daubert standard for legal evidence in that there is a known error rate.[7] The question that remains is whether this error rate can be further reduced by improved study designs. For example, should income and substitute reminders, cheap talk, and uncertainty scales all be applied concurrently in a study design, or is there overlap in what

[7] *Daubert v. Merrell Dow Pharmaceuticals, Inc.* 1993. 509 US 579, 113 S. Ct. 2786, 125 L. Ed. 2d 469 – U.S. Supreme Court. http://scholar.google.com/scholar_case?case=827109112258472814&q=daubert&hl=en&as_sdt=80000000000003, accessed June 9, 2010.

they accomplish, or do they have perverse effects that are not directly linked to providing accurate responses to stated preference questions? In addition, Foster and Mourato (2004) and Taylor, Morrison, and Boyle (2010) provide evidence that the overestimation error in choice experiments, which facilitate demonstration of sensitivity to scope, may be larger than what has been observed on contingent valuation studies and experimental settings. While the format of contingent valuation questions has been reduced primarily to variants of dichotomous-choice questions, convergent validity has not been conclusively established, nor has one response format been demonstrated conclusively as the best approach (Bateman et al., 2001, 2009). In addition, questions remain about the sensitivity of valuation estimates to the design of choice experiments (Özdemir and Boyle, 2009; Siikamäki and Layton, 2007).

In sum, the research undertaken in this monograph addressed fundamental questions in the application of contingent valuation at the time of the *Exxon Valdez* oil spill, and the outcomes of this research motivated fundamental changes in the processes by which stated-preference studies are implemented today. While the bounds of investigation on the validity of stated-preference methods have been refined substantially over the past 23 years, the fundamental questions posed in the research reported here still exist in some variant for today's applications of stated-preference methods.

References

Alberini, Anna, and Kevin J. Boyle. 2003. "Analysis of Contingent Valuation Data with Multiple Bids and Response Options Allowing Respondents to Express Uncertainty." *Journal of Environmental Economics and Management*, Vol. 45 (1): 40-62.

Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, and Howard Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." *Federal Register*, Vol. 58 (10): 4601-4614.

Bateman, Ian J., Brett H. Day, Diane P. Dupont, and Stavros Georgiou. 2009. "Procedural Invariance Testing of the One-and-One-Half-Bound Dichotomous Choice Elicitation Method." *The Review of Economics and Statistics*, Vol. 91 (4): 806-820.

Bateman, Ian J., Ian H. Langford, Andrew P. Jones, and Geoffrey N. Kerr. 2001. "Bound and Path Effects in Double and Triple Bounded Dichotomous Choice Contingent Valuation," *Resource and Energy Economics*, Vol. 23 (3): 191–213.

Boyle, Kevin J., and Richard C. Bishop. 1988. "Welfare Measurements Using Contingent Valuation: A Comparison of Techniques." *American Journal of Agricultural Economics*, Vol. 70 (1): 20-28.

Boyle, Kevin J., Richard C. Bishop, and Michael P. Welsh. 1985. "Starting Point Bias in Contingent Valuation Bidding Games." *Land Economics*, Vol. 61 (2): 188-194.

Boyle, Kevin J., William H. Desvousges, F. Reed Johnson, Richard W. Dunford, and Sara P. Hudson. 1994. "An Investigation of Part-Whole Biases in Contingent-Valuation Studies." *Journal of Environmental Economics and Management*, Vol. 27 (1):64-83.

Boyle, Kevin J., F. Reed Johnson, and Daniel W. McCollum. 1997. "Anchoring and Adjustment in Single-Bounded, Contingent-Valuation Questions." *American Journal of Agricultural Economics*, Vol.79 (5): 1495-1500.

Boyle, Kevin J., F. Reed Johnson, Daniel W. McCollum, William H. Desvousges, Richard W. Dunford, and Sara P. Hudson. 1996. "Valuing Public Goods: Discrete versus Continuous Contingent-Valuation Responses." *Land Economics*, Vol. 72 (3): 381-396.

Champ, Patricia A., Richard C. Bishop, Thomas C. Brown, and Daniel W. McCollum. 1997. "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods." *Journal of Environmental Economics and Management*, Vol. 33 (2): 151-162.

Cummings, R. G., D. S. Brookshire, and W. D. Schulze. 1986. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*. Rowman and Allanheld Publishers: Totowa, NJ.

Cummings, Ronald G., and Laura O. Taylor. 1999. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review*, Vol. 89 (3): 649-665.

Davis, Robert K. 1963. "Recreation Planning as an Economic Problem." *Natural Resources Journal*, Vol. 3 (2): 239-249.

Desvousges, William H., V. Kerry Smith, and Matthew P. McGivney. 1983. *A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvements*. Report to the U.S. Environmental Protection Agency.

Desvousges, William H., V. Kerry Smith, and Ann Fisher. 1987. "Option Price Estimates for Water Quality Improvements: A Contingent Valuation Study for the Monongahela River." *Journal of Environmental Economics and Management*, Vol. 14 (3): 248-267.

Dickie, Mark, Ann Fisher, and Shelby Gerking. 1987. "Market Transactions and Hypothetical Demand Data: A Comparative Study." *Journal of the American Statistical Association*, Vol. 82 (397): 69-75.

Foster, Vivien, and Susan Mourato. 2003. "Elicitation Format and Sensitivity to Scope." *Environmental and Resource Economics*, Vol. 24 (2): 141-160.

Haab, Timothy C., and Kenneth E. McConnell. 1997. "Referendum Models and Negative Willingness to Pay: Alternative Solutions." *Journal of Environmental Economics and Management*, Vol. 32 (2): 251-270.

Heberlein, Thomas A., and Richard C. Bishop. 1986. "Assessing the Validity of Contingent Valuation: Three field experiments." *The Science of the Total Environment*, Vol. 56: 99-107.

Heberlein, Thomas A., Matthew A. Wilson, Richard C. Bishop, and Nora Cate Schaeffer. 2005. "Rethinking the Scope Test as a Criterion Validity in Contingent Valuation." *Journal of Environmental Economics and Management*, 50 (1): 1-22.

Holmes, Thomas P., and Wiktor L. Adamowicz. 2003. "Attribute-Based Methods." Chapter 6 in *A Primer on Nonmarket Valuation*, P.A. Champ, K.J. Boyle, and T.C. Brown (eds.). Kluwer Academic Publishers.

Kahneman, Daniel. 1986. "Comments on the Contingent Valuation Method." In *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*, R.G. Cummings, D.S. Brookshire, and W.D. Schulze (eds.). Rowman and Allanheld Publishers: Totowa, NJ.

Kotchen, Matthew J., and Stephen D. Reiling. 1999. "Do Reminders of Substitutes and Budget Constraints Influence Contingent Valuation Estimates? Another Comment." *Land Economics*, Vol. 75 (3): 478-482.

Loomis, John, Armando Gonzalez-Caban, and Robin Gregory. 1994. "Do Reminders of Substitutes and Budget Constraints Influence Contingent Valuation Estimates?" *Land Economics*, Vol. 70 (4): 499-506.

Mitchell, Robert C., and Richard T. Carson. 1981. *An Experiment in Determining Willingness to Pay for National Water Quality Improvements*. Report to the U.S. Environmental Protection Agency.

Mitchell, Robert C., and Richard T. Carson. 1993. "The Value of Clean Water: The Public's Willingness to Pay for Boatable, Fishable, and Swimmable Quality Water." *Water Resources Research*, Vol. 29 (7): 2445-2454.

Murphy, James J., P. Geoffrey Allen, Thomas H. Stevens, and Darryl Weatherhead. 2005. "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation." *Environmental and Resource Economics*, Vol. 30 (3): 313-325.

Özdemir, Semra, and Kevin J. Boyle. 2009. "Convergent Validity of Attribute-Based, Choice Questions in Stated-Preference Studies." *Environmental and Resource Economics*, Vol. 42 (7): 247-264.

Samples, Karl C. 1985. "A Note on the Existence of Starting Point Bias in Iterative Bidding Games." *Western Journal of Agricultural Economics*, Vol. 10 (1): 32-40.

Siikamäki, Juha, and David F. Layton. 2007. "Discrete Choice Survey Experiments: A Comparison Using Flexible Methods." *Journal of Environmental Economics and Management*, Vol. 53 (1): 122-139.

Taylor, Laura O., Mark D. Morrison and Kevin J. Boyle. 2010. "Exchange Rules and the Incentive Compatibility of Choice Experiments." *Environmental and Resource Economics* (forthcoming).

Thayer, Mark A. 1981. "Contingent Valuation for Assessing Environmental Impacts: Further Evidence." *Journal of Environmental Economics and Management,* Vol. 8 (1): 27-44.

Veisten, Knut, Hans Redrik Hoen, Stale Navrud, and Jon Strand. 2004. "Scope Insensitivity in Contingent Valuation of Complex Environmental Amenities." *Journal of Environmental Economics and Management*, Vol. 73 (4):317-331.

Preface to the Second Edition

William H. Desvousges

More than 20 years ago, in March of 1989, the tanker *Exxon Valdez* went aground in Prince William Sound in Alaska. Although this event has had a profound effect on many different aspects of environmental affairs, I think few appreciate the significance it had on the field of natural resource damage assessment (NRDA). During my tenure at the Research Triangle Institute (RTI), I had co-authored the economics handbook for the U.S. Department of the Interior (DOI) to accompany the NRDA regulations that were promulgated in 1986. After writing that handbook, Rick Dunford and I wrote an appraisal of the economic issues that were likely to be significant in implementing the NRDA regulations for the American Petroleum Institute (API). We were able to reflect some of the experience that we had gained working on the 1988 Shell oil spill in Martinez, California, which was the first significant damage assessment to occur after the promulgation of the DOI regulations. We had sent the final draft of our report to API in January, just a few months before the *Exxon Valdez* went aground. We both were wondering whether anything was going to come of this new field of damage assessment, or whether it was going to pass into the night without having much of an effect on either our lives or environmental economics.

Little did we know what was in store when API gave a copy of our report to Gary Dowling and Diane Schenke of Exxon, who had come to Washington to deliver a check for $15 million to the federal government to fund the assessment of the environmental consequences of the spill. Within a couple of weeks, Rick and I and RTI had been retained by Exxon to assist them with the environmental economic issues related to the spill. We devoted a large share of our professional lives for the next 3 years working on this endeavor. Rick and I completed a wide range of tasks as part of this research, including several trips to Alaska and interacting with two Nobel Prize–winning economists, Kenneth Arrow and George Stigler, on the economic issues. (A third member of the Exxon economics team, Dan McFadden, would subsequently win a Nobel Prize in economics.) This monograph is one result of the research we performed for Exxon, and the one that has endured more than any other.

By the time we began working on the research contained in this monograph, Reed Johnson had joined the team, first on a part-time basis while retaining his position at the Naval Academy and then joining RTI full-time in 1990. We also were fortunate that Kevin Boyle, who was at the University of Maine at that time, was on a sabbatical for a year at North Carolina State University. Kevin also had been working for Exxon after having been involved in a little-known trade of economic experts. After the spill, Kevin had been retained by attorneys working for the government and Gardner Brown of the University of Washington had been retained by outside counsel working for Exxon. The two professors were interested in working for the other side, so the parties agreed to release each from their contracts and allow them to pursue work for the other party. Thus, we were able to obtain Kevin's expertise during this critical time. Sara Hudson had joined RTI and was working with Reed, Rick, and me as part of the overall project. Sara played several invaluable roles in completing both the underlying research and the monograph itself.

With this team in place and working on behalf of Exxon, we responded to a request for proposal (RFP) that Exxon sent to the various economists who had been retained to work on the damage assessment. The RFP contained a list of ideas on which Exxon was interested in undertaking research to address various aspects of measuring so-called nonuse values, which are independent of any use of natural resources. By that time, everyone had agreed that this topic was likely to be the focal point of any debate on environmental damages between Exxon and both the federal government and the State of Alaska. We responded to several of the items in this list. Our first proposal was designed to address a fundamental question: can the survey methods that are used to measure nonuse values accurately measure changes in the quantity of the good being valued in the survey? Our proposal, which later formed the basis for the "birds study," was stimulated by an article that Rick had seen in the *News and Observer* that indicated more birds had died by landing in waste oil holding ponds in the Central Flyway than had been killed in the *Exxon Valdez* oil spill. We decided that this would provide a simple to explain context that was less likely to be as emotionally laden as developing a survey related to the spill itself. At the time, I thought that this study would be one in which we would develop the questionnaire, administer it, and show that the survey would pass such a simple test.

I thought that the companion study in this monograph that compares valuations to prevent large oil spills and small oil spills would be the study that would attract the most attention by the various parties in the debate. Once again, my ability to predict the future would prove to be woefully inaccurate.

The simple premise of the birds study was whether you would get the same valuation answer in a survey for protecting 2,000, 20,000, or 200,000 birds. We designed different versions of the survey questionnaire and administered it to three separate samples of respondents. We found, much to my surprise, that the respondents gave essentially the same answer for protecting the three very different numbers of birds. Our simple intuitive results resonated with many participants in the damage assessment debate, including the blue-ribbon panel of experts that had been formed by the National Oceanic and Atmospheric Administration (NOAA) in 1991 to address the reliability of the use of survey methods to measure nonuse values. Our birds study was one of the few studies mentioned in the subsequent report issued by the panel in 1993. The birds study also figured prominently in the NOAA panel's requiring that any survey study that was to be used to measure nonuse values would have to pass the so-called scope test, which requires that independent samples be used to determine whether survey valuation responses change with increases in the scope of the injury being described in the survey. Although various debates have ensued about the accuracy of our results, no one can deny that the requirement of the scope test fundamentally changed the way economists considered and designed surveys to measure nonuse values. Subsequently, we published an article in the *Journal of Environmental Economics and Management* that provided a more complete discussion of the issues, but the basic underpinnings of the research are described here in this monograph. The oil spill experiment has gone virtually unnoticed, even though it produced similar results.

Both Rick and I continue to labor in the fields created by the research conducted in response to the *Exxon Valdez* more than 20 years after the spill. We have worked on NRDAs throughout the United States—and beyond, in Rick's case. This research and subsequent studies led Rick, Reed, and me to form our own firm, Triangle Economics Research, to continue the investigations. We were fortunate that RTI provided us the support to be able create our firm and maintain good relationships. Now, as more time has passed, Reed has returned to RTI to work on cutting-edge valuation topics in health economics, while Rick and I have formed small companies that allow us to continue investigating economic topics that stemmed from the *Exxon Valdez* research. Kevin also continues to conduct research on these topics, and we were pleased that Kevin was willing to write the Forward for this reprinted edition of the monograph to provide his perspective on the state of things. Sara, who has shown perhaps the most sense of all of us, is now teaching school where her daughters had attended. With her intellectual acumen and her communication skills, we know that the children who pass through her classroom are indeed the luckiest ones.

This monograph would not have been possible without the commitment of Exxon through its completion. Mike Noland, Mike Denning, and Robert Hirsch worked closely with us throughout the various stages of the research. They provided invaluable guidance and comments on all aspects of the research. Gary Dowling was a valued supporter throughout our time working for Exxon. George Lock supervised the overall direction of the damage assessment program after the first year of our work with Exxon and oversaw the study's completion.

Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy

William H. Desvousges
F. Reed Johnson
Richard W. Dunford
Kevin J. Boyle
Sara P. Hudson
K. Nicole Wilson

RTI

Preface to the First Edition

This monograph provides the results of some experiments that were funded by Exxon as part of its research associated with the 1989 *Exxon Valdez* oil spill in Alaska. Exxon decided to undertake these experiments to test the accuracy of contingent valuation for measuring nonuse values. However, the results reflect the opinions of the authors and not necessarily those of Exxon. Also, the opinions expressed do not necessarily reflect those of Research Triangle Institute.

Bill Desvousges presented some of the results reported in this monograph at the Cambridge Economics Inc. symposium titled “Contingent Valuation: A Critical Assessment” on April 2-3, 1992, in Washington, DC. This volume is intended to provide more details about the experiments for those interested in such details. The monograph format also allows us to expand our discussion of many topics beyond the scope of a single paper or a series of papers.

In preparing this volume, we have benefitted from many discussions with other researchers working for Exxon. We would especially like to thank Jerry Hausman, Dan McFadden, John Payne, and David Schkade.

Many RTI staff contributed to the research reported in this monograph. We would like to thank Gay Todd Shackelford, A. Jeanne Milliken, and Kristin J. Stettler for their contributions to the development and testing of the questionnaires. Also, we especially want to thank Andrew Jessup, Judy Parsons, and Judy King for their diligence, patience, and overall excellent work in the production of this monograph.

Executive Summary

Research Triangle Institute (RTI) conducted a series of experiments designed to evaluate the accuracy of contingent valuation for measuring natural resource damages. This monograph tests hypotheses related to three basic elements of accuracy:

- **Theoretical Validity**, which is concerned with the extent to which estimates are consistent with theoretically derived hypotheses.
- **Convergent Validity**, which requires that estimates for a given commodity be the same, regardless of the elicitation framework used.
- **Reliability**, which requires estimates to be robust to variations in the statistical analyses.

This monograph presents the results of two independent experiments that evaluate these elements of CV's accuracy for measuring nonuse damages. Both experiments are intended to measure the total value that people place on specific resource services, which is consistent with the proposed DOI regulations. The results of the experiments indicate that CV is unable to measure these total values accurately in a natural resource damage assessment context.

The first experiment examines whether hypothetical willingness-to-pay (WTP) responses to contingent valuation questions increase as the level of migratory waterfowl protection increases. Specifically, our experiment measures WTP for three levels of migratory waterfowl protection: 2,000, 20,000, and 200,000 deaths prevented annually. These levels of bird deaths span the levels that occur in actual oil spills. The level of bird deaths avoided is the only difference in the three versions of the questionnaire. The hypothetical scenario involves regulations requiring covers for waste-oil holding ponds in the southwestern United States that would prevent migratory waterfowl in the Central Flyway from landing in the ponds.

The second experiment elicits hypothetical WTP responses for two levels of oil-spill response to reduce the environmental effects of spills: reducing the environmental effects of small spills only and environmental reducing the effects of all spills. The experiment includes the exact wording of the small-spills version within the all-spills version and then describes the additional response measures that would be provided for large spills. This experiment proposes a national policy that would be in addition to current regulations. Additionally, in this experiment we compare two question formats that are widely used in the CV literature: the open-ended format and the dichotomous-choice format.

For these experiments, we developed questionnaires using state-of-the-art procedures, including focus groups, one-on-one interviews, and large pretests. Self-administered questionnaires were obtained from 1,205 respondents in the migratory waterfowl experiment and 1,607 respondents in the oil-spill experiment. We randomly chose non-student respondents 19 years or older at two shopping malls in suburban Atlanta, Georgia. We chose the location for the experiment and defined the commodities so that nonuse values would be the dominant element.

The principal objective of these experiments is to test three basic hypotheses:

- **Hypothesis 1**: WTP estimates increase for higher levels of natural resource services. This hypothesis is a test of theoretical validity.
- **Hypothesis 2**: The open-ended and dichotomous-choice CV formats yield comparable WTP estimates for the same natural resource services. This hypothesis is a test of CV.
- **Hypothesis 3**: The dichotomous-choice WTP estimates are not sensitive to reasonable choices in the analysis of the data. This hypothesis is a test of reliability.

To test these hypotheses, we perform both nonparametric and parametric tests. Rejection of the null hypotheses implies that the tested accuracy properties are not established in our experiment.

Our results reject Hypothesis 1. We find that the WTP estimates do not increase as the level of natural resource services increases. Specifically, WTP does not increase for the three levels of waterfowl protection or the two levels of oil-spill response. Furthermore, the WTP estimates for the open-ended small-spills version exceed those for the open-ended all-spills version. The results are confirmed in statistical tests of the entire distributions of responses, as well as in tests comparing measures of central tendency such as means. They are also confirmed in tests using more sophisticated models to explain responses.

Additionally, the oil-spill results hold for both the open-ended and dichotomous-choice question formats. There is no increase in WTP from the small-spills to all-spills version based on either tests of the distributions or the estimated means. This latter finding is significant because some CV proponents argue that the dichotomous-choice format is the preferred format.

For Hypothesis 2, our results also show that the two question formats (open ended and dichotomous choice) do not produce comparable estimates of WTP for oil-spill protection. We use two approaches to test Hypothesis 2. First, we developed a dichotomous-choice data set from the open-ended WTP responses to the oil-spills questionnaires by randomly assigning one of the six dichotomous-choice bids to each open-ended response. Assuming that respondents would have accepted the bid if their stated WTP was at least as large as the randomly assigned bid, we produced a synthetic set of dichotomous-choice responses. We then compare the results from this synthetic dichotomous-choice data set with the results from the actual dichotomous-choice data. Second, we directly compare the means estimated using the dichotomous-choice data with the corresponding means estimated with the open-ended data. Using these two approaches, we find significant differences in the distributions of responses for the two formats. These differences are caused largely by the implausible occurrence that 34 percent of the respondents agreed to pay the offered $1,000 bid in the dichotomous-choice survey, while only 3 percent of respondents in the open-ended survey indicated a WTP greater than or equal to $1,000. The majority of our means tests also confirm the differences between the question formats. The lack of comparability is important because neither format can be judged superior on its own merits, nor is the true WTP known for determining which format is best.

For Hypothesis 3, we look at the effect of alternative estimation strategies on the dichotomous-choice WTP estimates. Specifically, we investigate the effect of alternative functional forms, limits of integration, and bid structures. We show that the dichotomous-choice method is very sensitive to these choices, all of which are subject to the analyst's discretion. We conclude that two analysts using a given data set with standard techniques could produce results that differ by several orders of magnitude in the most extreme case.

Finally, we conclude based on our experiments that CV estimates of nonuse values fail to reflect large and realistic differences in levels of natural resource services. Although our study does not determine the cause of this failure, the policy implications are clear. CV estimates of nonuse damages are neither valid nor reliable enough to be included in natural resource damage assessments where responsible parties are required to write a check for fair compensation for the loss in natural resource services.

CHAPTER 1

Background

1.1 Introduction

In 1986, the U.S. Department of the Interior (DOI) promulgated regulations for conducting natural resource damage assessments (NRDAs) under the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) of 1980 (see Desvousges, Dunford and Domanico, 1989). These regulations provide protocols for conducting assessments, as well as delineating appropriate measures of compensation for injuries to natural resources resulting from releases of oil or hazardous substances. The regulations focus on the residual injury that may remain after clean-up or remediation activities are completed. Originally, the regulations prescribed that Trustees use the lesser of restoration costs or foregone use values[1] as the basis of measuring damages. They allowed for nonuse values[2] only when no use values could be measured.

In the *State of Ohio v. Interior* decision in 1989, the Court of Appeals broadened the scope of nonuse values, or "passive use" in their lexicon, to all situations where these values can be reliably measured (see Dunford, 1992; and Carson, Hanemann, and Kopp, 1991). The Court did not define reliability, however. In response to the Court's ruling, the DOI is currently revising the NRDA regulations (see 56 Fed. Reg. 19752-19773; and Desvousges and Dunford, 1991). The proposed

---

[1] Use values are values associated with uses of natural resources such as sport fishing, boating, and hunting.

[2] We define "nonuse values" as any values not directly associated with human uses of natural resources. This definition is consistent with the NRDA regulations.

changes to the NRDA regulations include nonuse values in the evaluation of restoration alternatives, compensable value,[3] and preliminary estimates of damages. Nevertheless, the views on measuring nonuse damages in the proposed changes in the NRDA regulations are still evolving. Comments on the proposed changes by economists who have worked on opposite sides in recent cases reveal interesting similarities.

> ...economists agree that the existence of close substitutes eliminates or, at the very least, minimizes the proportion of total value (or compensable value) attributable to nonuse values. Nonuse damages should be available only where the injured resource is unique and injured irreversibly (Cicchetti, 1991).
>
> The magnitude of nonuse values is a purely empirical matter which will vary from case to case. It is certainly affected by the extent of the injury and, also, the presence of substitutes (Carson, Hanemann, and Kopp, 1991).
>
> As mentioned in the preamble, ...the DOI should limit the inclusion of nonuse losses in damage assessments to situations where the Trustees can demonstrate irreversible or long-lasting changes to unique, widely-recognized natural resources (Desvousges and Dunford, 1991).

Each shows subtle differences in emphasis on the role of substitutes in determining the size of nonuse damages. As noted by Carson, et al. (1991), the magnitude of these damages is an empirical issue, and empirical evidence is needed to address nonuse values. In this report, we provide some evidence on the accuracy of contingent valuation (CV) in measuring nonuse values for natural resource damage assessments.

Recent government regulations on oil spills have increased the importance of measuring nonuse values. In 1990, the U.S. Congress passed the Oil Pollution Act (OPA) to reduce both the chances of oil spills occurring and the effects of spills on the environment. The OPA requires oil companies to

- build new tankers with double hulls and add double hulls to existing tankers,
- pay higher fines and penalties for spilling oil,

[3] The proposed changes in the DOI regulations define compensable value as "all the public economic values associated with an injured resource, including use values and nonuse values ..." (56 Fed. Reg. 19760).

- upgrade their navigational systems, and
- provide more funds for oil-spill research.

Although the requirements outlined in the OPA primarily are focused on prevention, the OPA also stipulates that oil companies must pay damages for losses in natural resource services that include reliably measured nonuse values, along with paying higher fines and penalties for spilling oil. The OPA, coupled with CERCLA and the *Ohio v. Interior* ruling, indicate the increasing awareness and political concern for nonuse values as a component of natural resource damages. Moreover, state and federal governments have sought substantial damages from potentially responsible parties on the basis of lost nonuse value.[4] Therefore, determining whether damages for losses in nonuse values can be reliably measured is a crucial concern in NRDA cases.

1.2 Overview of Nonuse CV Literature

The CV approach uses surveys to directly measure people's willingness to pay (WTP) for nonmarket goods. Although CV traces its lineage to a study by Davis (1964) of hunting in the Maine woods, most current applications stem from the work of Randall, Ives, and Eastman (1974) and the various studies performed at the University of Wyoming (see d'Arge, Schulze, and Brookshire, 1980; Rowe, d'Arge, and Brookshire, 1980; and Schulze, d'Arge, and Brookshire, 1981; as examples).

The early CV studies focused on developing empirical estimates of theoretical welfare measures. They particularly concentrated on measuring the benefits of air quality regulations, although various methodological experiments also were undertaken (see Brookshire, Randall, and Stoll, 1980; Brookshire, Eubanks, and Randall, 1983; and Thayer, 1981). Walsh, Loomis, and Gillman (1984) also applied CV to address a broad group of natural resource policy questions such as wilderness preservation, stream flow preservation, and water quality enhancements. (In addition, see Greenley, Walsh, and Young, 1981; and Daubert and Young, 1981.) In

[4] The natural resource damages from the Shell oil spill near Martinez, California, in 1988 were settled for approximately $11 million. The State of Washington and the province of British Columbia alleged natural resource damages in excess of $10 million for the *Nestucca* oil spill. The New Bedford Harbor cases have involved claims for natural resource damages that exceed $20 million.

the early 1980s, Mitchell and Carson (1981 and 1984) and Desvousges, Smith, and McGivney (1983) extended the methodology by enhancing questionnaire design techniques and using more rigorous sampling and survey procedures.

Most of the early studies focused on use values such as health benefits, visibility, and water quality improvements. Recently, the focus of several CV studies has turned to measuring nonuse values, especially existence values[5] that do not require use of the resource. Economists' experience in measuring nonuse values is quite limited compared with their experience in measuring use values. Of the approximately 170 empirical CV studies cited in Mitchell and Carson (1989), only about 20 address commodities for which nonuse values are a substantial portion of total value. Appendix A summarizes 18 of the most important studies of total or nonuse values that have appeared in the last 11 years. Table 1-1 summarizes some common characteristics of these studies. Over two-thirds of these studies estimate nonuse values either directly or indirectly. Sample sizes for these studies range from 140 to 2,561 and they use many different question formats and payment vehicles.

Table 1-1. Characteristics of 18 Recent Total/Nonuse Value Empirical Studies

Estimated value	
Total value only	5
Nonuse value derived from total value	10
Nonuse value directly	3
Type of interview[a]	
Mail	10
Personal	9
Sample size[b]	
<500	6
500–1,000	11
>1,000	2
Question format	
Open-ended	5
Dichotomous choice	5
Bidding game	2
Payment card	2
Multiple formats	4

[a] One study used both mailed questionnaires and in-person interviews.
[b] One study had two separate mailed questionnaires.

These studies generally have focused on permanent changes in unique resources, usually in an *ex ante* setting that incorporates alternative regulatory or management scenarios (e.g., developing a dam versus preserving a free-flowing river or attaining

[5] The NRDA regulations define existence value as the value of "individuals who do not plan to utilize a resource now or in the future, but are willing to pay to know that the resource would continue to exist in a certain state of being" (51 Fed. Reg. 27692).

alternative levels of water-quality improvement for a river basin). For measuring natural resource damages from oil spills, the key damage measurement issues involve temporary changes in the level of services from resources that may range from ordinary to unique. The pressing empirical issue is whether CV is capable of reliably measuring the kinds of changes in the services of natural resources that are relevant for oil spills.

Despite the relative inexperience in estimating nonuse values, most CV researchers appear confident that methods developed for use values are equally valid for nonuse values. For example, Mitchell and Carson (1989) state that CV "can obtain valid benefit measures of amenities which include an existence component, ...provided the potential problems are recognized and overcome." In response to the *Ohio v. Interior* (1989) ruling, Carson, Hanemann, and Kopp (1991) assert that all of the valuation techniques mentioned in the proposed NRDA rule, including CV, are "capable of giving accurate and reliable estimates if there are good baseline data, good scientific information, experienced researchers, and adequate resources of time and money," but "CV is the *only* method available for the measurement of nonuse values."

This confidence would be justified if nonuse values posed no new problems for researchers or respondents compared with use values. However, some critics of using CV to measure nonuse values argue that such estimates are more susceptible to well-known framing and incentive bias problems as well as highly variable WTP results (see Kahneman and Knetsch, 1992; and Phillips and Zeckhauser, 1989). Recent research suggests that such problems arise at least partially because of respondents' lack of familiarity with nonuse services and lack of experience in valuing such services (Schkade and Payne, 1992; Milgrom, 1992). The current debate has been especially lively and sometimes heated following the Kahneman and Knetsch (1992) paper. Part of the response stems from concerns over whether CV critics actually have demonstrated their main points (see Smith's [1992] comments on Kahneman and Knetsch). The debate also is fueled by the magnitudes of potential liability for nonuse damages.

Because nonuse values may be assessed against potentially responsible parties for oil and hazardous-substance releases, much of the current research is pursued in the highly charged atmosphere of litigation. Although this context increases the urgency of validating values for which someone actually will write a check, it also inhibits the peer review that helps establish the scientific credibility of research results.

1.3 Accuracy and NRDA

The NRDA process is another step on the evolutionary path of benefit-cost analysis from the Green Book, to the Water Resources Council Guidelines, to regulatory impact analyses under Executive Order 12291. Many of the same benefit-cost principles apply and much of the past experience is relevant to damage assessments. The basic rationale of damage assessments is consistent with the applied welfare principles underlying benefit-cost analyses. The fundamental reason for requiring those responsible for injuries to natural resources to pay compensation is to protect the value of the national portfolio of natural assets. This goal is best achieved when impaired resource services are restored efficiently and when firms face appropriate incentives to avoid or mitigate future injuries to resource services.[6]

The need to send appropriate signals to firms and the need for fair compensation place a premium on accurate value measures in NRDAs. Accuracy consists of two important dimensions: validity and reliability. Validity primarily is the extent to which a measurement technique produces an unbiased estimate of the true value it is designed to measure. Other aspects of validity also include whether measures conform to theoretical expectations and whether measures across different designs, such as different ways to elicit values, converge to a single value. Reliability concerns the extent to which variation in experiment results over repeated trials falls within an acceptable error range and the extent to which statistical results are sensitive to issues such as functional form (see Carmines and Zeller, 1979; Rossi, Wright, and Anderson, 1983; and Allen and Yen, 1979).

[6] If the "lesser of" provision in the original DOI regulations had specified total values, then it would have come closer to the economically efficient outcome, assuming that total values can be measured reliably. Responsible parties would have paid damages equal to the lesser of foregone total values or restoration costs. When restoration costs exceeded total values, then natural recovery was the logical choice of action and payment was made for the full value of the foregone services. When it was less expensive to restore services than pay the foregone total value, then restoration was the appropriate choice. The Court of Appeals decision requiring restoration in all cases limits society's flexibility in choosing the best course of action.

We think that CV analysts have given too little attention to the function that CV estimates must perform in NRDAs. As Smith notes in his commentary in Cummings, Brookshire, and Schulze (1986):

> One reason why there has been diversity among CVM researchers in their judgments as to its performance is the use they intend for the benefits estimates. ...We may be able to tolerate low levels of accuracy for some purposes. It appears that those evaluating CVM have quite different end users in mind. The old adage—"good enough for government work"—may well be literally relevant in some applications of CVM estimates. ...By contrast, tests of specific hypotheses, or indeed some benefit-cost decisions, may hinge on the accuracy of the estimates of individual valuation....CVM may prove acceptable in some cases and not others. (p. 175)

Certain features of damage assessments raise fundamental concerns about the accuracy of damage measures. In particular, six important features of damage assessments should be considered in evaluating accuracy. Each of these is discussed below.

First, use values, which have received considerable research attention in past studies, seldom are sources of controversy in most assessments. We have conducted several oil-spill assessments over the past few years, including the Shell spill in northern California and the Exxon spill in the Arthur Kill between New Jersey and New York. In both of these situations, the responsible party and the Trustees had little disagreement about foregone use values. (In the northern California case, which involved an area with a high level of recreation use, foregone use values were a significant part of the assessment.) In other situations, use value losses may be more of an issue. For example, situations with difficult baseline determinations and long contamination periods could involve substantial disagreements about foregone use values. However, these disagreements are likely to involve parameters such as the price elasticity of demand and the timing of the losses.

Second, nonuse values will be the primary focus of controversy. These values have received much less attention in past studies. Research funded by the Environmental Protection Agency and the Forest Service stimulated many of the early CV investigations and the research on nonuse values. Debate in the early 1980s over the efficacy of the research culminated in the Palo Alto Workshop, which was ultimately reported in the Cummings, Brookshire, and Schulze (1986) volume.

One interpretation from the workshop is that CV is likely to be least reliable when it is most needed—that is, for measuring nonuse values for which other valuation methods do not exist. However, as Smith and Desvousges (1986b) note, this conclusion is not based on systematic research. Ironically, budget cuts and changing priorities resulted in little subsequent research to investigate the accuracy of CV for measuring nonuse values. Consequently, much of the experience with CV stems from use-value studies and from some studies that have addressed total value. The relevance of the use-value experience for measuring nonuse values remains a central question.

Third, the extent of the market, especially for nonuse services, has received little empirical attention. The size of the market can significantly affect the magnitude of damages, because CV tends to yield average household or individual values that are multiplied by some number of households or individuals. Few studies have addressed the fundamental definition of the market for nonuse services. (Smith [1992b] is a notable exception.) Since these services do not require actual use of the resource, the market for these services could potentially include the entire country. Yet, intuition would suggest that the size of the market would vary according to the characteristics of the resource. We would expect that the market for common resources would be smaller than the market for unique resources such as the Grand Canyon. A recent spill offers an interesting illustration of the importance of the relevant market. The *Nestucca* spill affected shoreline in both the U.S. and Canada. The main resource service affected by the spill was the loss of a disputed number (4,000 to 40,000) of common murres. The interaction of a relatively marginal change in the stock of murres with the long area of affected shoreline that includes two countries poses a challenge for trying to determine the relevant market. The marginal change in the affected resources would argue for a narrowly defined market area, but no empirical evidence is available at this time to support such a determination. Thus, while the concept of nonuse values does not impose any geographic limits, the nature of the services that are relevant for damage assessments will likely lead to more narrowly defined markets. How much narrower is the question that could have significant implications for the parties involved in NRDAs.

Fourth, the nature of damages in an NRDA often are temporary changes in the quality of resource services, not the permanent loss of services. Most nonuse studies have concentrated on long-term changes in quality or on the permanent loss of some natural resource. Under the proposed NRDA regulations, the only relevant damage category that CV can be used to measure is the value of the services foregone prior to restoration. For temporary changes in quality, the valuation task in CV may become more difficult for these interim services. Survey respondents will have to separate the value of the loss in some service flows for temporary periods from the value that they place on the asset itself. The CV instrument must be able to communicate to people the types of restoration activities that are undertaken and the time it will take for the resource to return to its without-spill condition. In addition, scientists and economists will have to explain these service flows in terms that people can understand, but will still correspond to the services that are actually affected by the oil spill. The uncertainty associated with injuries and recovery periods further complicates an already difficult valuation task. No research to date has tested whether nonuse values for temporary losses in quality exist or, if they exist, what factors limit their magnitude.

Fifth, an oil-spill setting often is highly emotional, accompanied by the dissemination of contradictory and frequently inaccurate information from various sources. This information is likely to affect any attempt to measure nonuse values with CV. Even though natural resource damages are supposed to be compensatory, punitive motives may also influence the WTP responses of some CV respondents in emotional oil-spill settings. At this stage in the research on NRDAs, no one has clearly stated which assumptions and requirements the information must satisfy to obtain acceptable measures of nonuse values that exclude punitive motives.

Finally, one firm or a small group of firms will have to write a check for the assessed amount of natural resource damages. However, if nonuse services were affected by the spill, the amount of this check will be based partially or largely on hypothetical values elicited in a CV survey. To the extent that these hypothetical values overstate respondents' actual values for the foregone natural resource services, then responsible parties will pay more in natural resource damages than socially desirable. To date, no published studies have compared respondents' hypothetical WTP for nonuse services in a CV survey to their actual WTP.

Therefore, accuracy in NRDAs is not merely an academic issue. To be a defensible basis for compensable damages, valuation methods must produce estimates of resource value that can be verified in repeated studies. Additionally, the estimates must be related to the specifics of the loss in services in order for compensation to reflect the actual loss, not merely general environmental anxiety. Finally, the damage estimates should reflect society's actual (not hypothetical) value for the affected services since responsible parties will actually write a check to Trustees for the estimated value of the service losses.

An evaluation of the accuracy of damage estimates must consider both validity and reliability. For example, although a high level of reliability is desirable, it does not indicate that a value is unbiased or valid. Simply arriving at the same answer in test-retest comparisons is not enough. Additional tests are needed to ensure that the answer is also valid. Finally, damage-assessment checks must be written on point estimates, not 90 percent confidence intervals. In this report, we explore several dimensions of accuracy as they relate to NRDAs.

1.4 Background on Experiments

The experiments described in this report are among the first fundamental tests of CV's measurement capabilities for nonuse losses. The first experiment tests whether people give higher CV responses for protecting larger numbers of migratory waterfowl. According to the U.S. Fish and Wildlife Service (USFWS) more than 100,000 migratory waterfowl die each year in the southwestern United States in uncovered ponds containing wastewater, oil, and other byproducts from oil and gas drilling operations. More than 250,000 of these ponds, which range from 10 to 100 feet in diameter and 14 to 100 feet in depth, are found in the Playa Lakes Region of eastern New Mexico, northwest Texas, western Oklahoma, and parts of Kansas and Colorado. Migratory waterfowl are attracted to these waste-oil holding ponds in the Central Flyway because there are few wetlands and freshwater ponds in the region. Affected migratory waterfowl include mallard ducks, pintail ducks, white-fronted geese, snow geese, and greater sandhill cranes (Hubert, 1990).

While the USFWS does not regulate or control oil and gas operations, it has authority under the Migratory Bird Treaty Act of 1918 to prosecute "wrongful" deaths of migratory waterfowl. Accordingly, in 1988 USFWS began encouraging states in the Southwest (mainly New Mexico, Oklahoma, and Texas) to require companies to cover their ponds. Since that time, New Mexico has been the most aggressive in requiring companies to cover their waste-oil holding ponds with a screen or netting (Kelly, 1990). Texas recently adopted a statewide rule requiring a screen, net, or cover on most waste-oil holding ponds (Conservation, 1991).

The survey questionnaire describes a regulation that would require wire-net covers on waste-oil holding ponds to protect migratory waterfowl. The questionnaires, which are identical in all other respects, indicate one of three levels of migratory waterfowl protection: 2,000, 20,000, and 200,000 bird deaths prevented annually. These three levels of waterfowl deaths reflect the range of bird deaths associated with many oil spills. For example, the Arthur Kill pipeline spill in January 1990 is said to have killed an estimated 1,200 to 2,000 birds, numbers very close to our 2,000-bird scenario.[7] The *Nestucca* spill in December 1988 is claimed to have killed an estimated 40,000 birds (see Rowe et al., 1991). For the *Exxon Valdez* spill, the Trustees estimate that 260,000 to 580,000 birds were killed (*Oil Spill Intelligence Report*, 1991b).

This test uses a commodity that is simple and concrete, which should provide a relatively easy situation for people to assess their preferences. It also involves a situation where respondents may have relatively few preconceived notions, which otherwise could complicate CV's ability to measure the value of specific natural resource services. Therefore, we expect that CV would work better in this experiment than in typical NRDAs, which often involve more emotional situations.

The second experiment tests whether WTP responses increase for an increased level of oil-spill response capability. Although the media coverage has focused on very large oil spills, most spills are relatively small in size. According to data collected by the Coast Guard for the years 1984 through 1986, well over 95 percent of all oil spills in U.S. waters were under 50,000 gallons in size. In fact, the majority of spills reported to the Coast Guard were under 500 gallons. During that 3-year period, the average size of a spill was 1,789 gallons (USDOT, 1989). The General Accounting Office, in a report to Congress, notes that about half of the oil spilled in 1988 was spilled near waterfront facilities, with a majority of that oil spilling during the transfer of oil between the facility and a vessel (GAO, 1991).

---

[7] This range is estimated using the number of dead birds found and a multiplier of three to five to account for unrecovered dead birds (*Oil Spill Intelligence Report*, 1991a).

The OPA of 1990 addresses some prevention issues, such as double-hulled tankers and improved navigational equipment. However, an increase in response capacity would help minimize the environmental damage from oil spills that still occur. After providing respondents with information on the OPA of 1990, oil spills, and the damages resulting from oil spills, the second experiment describes two types of response centers: (1) local response centers that would only prevent environmental damage from oil spills of less than 50,000 gallons and (2) regional response centers that would prevent environmental damages from large, offshore oil spills. The local response centers would be located at each port that handles oil. They would be equipped with inventories of response equipment (e.g., booms, skimmers) and local emergency response personnel would be trained in oil-spill response. Additionally, local volunteers would be trained in wildlife rescue. The regional centers would add to the capabilities of the local centers by preparing for larger spills. These regional centers would be located on the East Coast, Gulf Coast, West Coast, and in Alaska. They would employ full-time professional response staff and maintain larger inventories of response equipment. Also, they would provide facilities for large-scale wildlife rehabilitation. This experiment enables us to examine issues directly related to the context of oil spills, which is likely to be more emotional than the migratory-waterfowl experiment.

The oil-spill experiment also includes a test of whether the format of the WTP question (i.e., dichotomous-choice vs. open-ended) affects the magnitude of the WTP responses. Bishop and Heberlein (1979 and 1986) first used the dichotomous-choice, or referendum, format in their experimental work comparing CV with other valuation methods. The dichotomous-choice format has become the preferred question format for many CV practitioners. They argue that the "take it or leave it" style question, which only requires people to give a Yes or No response, more closely resembles a market choice than the open-ended alternatives. Hanemann (1984) also contributed to the adoption of the dichotomous-choice method by establishing a link between theoretical welfare concepts and empirical estimates of welfare gains and losses derived from the method. However, very little attention has been given to actually comparing question formats using independent samples. Most tests have compared responses for people who were given both formats (see Loomis, 1990;

Boyle and Bishop, 1988; Stevens et al., 1991; and Kealy, Dovidio, and Rockel, 1988).[8] These comparisons are of limited value because people's responses to one format are likely to condition their responses to the other. Our study employs independent samples for the comparison of question formats.

The choice of this overall design reflects several considerations:

- The migratory waterfowl scenario tests a commodity that is relatively simple and concrete.
- The commodity may have some use-value component, but nonuse values dominate the total value for most respondents. (The survey deals with the Central Flyway, and the interviews were conducted in Atlanta, Georgia.).
- The migratory waterfowl scenario has relatively little emotional content, which Freeman (1990) suggests should improve the ability of CV to accurately measure WTP.

These characteristics of the migratory waterfowl scenario are intended to minimize the likelihood that WTP values are proxies for general environmental attitudes. The design also measures total value, which Mitchell and Carson (1989) argue is the preferred way to structure the valuation task.

The oil-spill scenario addresses a broader set of issues that are relevant to many policy questions. It elicits an *ex ante* value for reducing the environmental effects of oil spills. The national policy context allows for a substantial nonuse component because it would protect all beaches, not just those the respondent might use. However, the oil-spills scenario lacks the simplicity of the migratory waterfowl scenario because oil spills sometimes are national, or even global, in significance. The wide media coverage received by large oil spills undoubtedly influences public perceptions. Preoccupation with culpability and punishment may make it more difficult to elicit respondents' WTP for resource services. The strong influence of

[8] Johnson, Bregenzer, and Shelby (1990) and Seller, Stoll, and Chavas (1985) use independent samples to compare question formats, but they do not conduct any formal hypothesis tests. Kriström (1988) is another exception, but this study remains unpublished and largely unavailable. He found that dichotomous-choice WTP estimates were substantially higher than open-ended WTP estimates for preserving virgin forests in Sweden.

information from outside the scenario also makes it more difficult to restrict the elicitation to the facts presented in the scenario. The migratory waterfowl scenario alone could be questioned regarding its relevance to policy decisions about oil spills. However, the oil-spill scenario allows us to check whether our results are replicable for another commodity. By performing two separate studies, we are able to test the consistency of our results. If independent studies produce similar results, the credibility of both studies is reinforced.

In summary, we perform two independent, yet related, CV surveys for our study. The test of question formats also uses independent samples. As discussed in Chapter 3, we used focus groups, one-on-one interviews, and large pretests to refine the survey instruments for both experiments. We have large sample sizes to strengthen the statistical analysis. We also employ some of the latest analysis techniques, such as bootstrapping, to test various hypotheses.

1.5 Organization of Monograph

Chapter 2 addresses accuracy and discusses its importance for estimating nonuse values using CV. Chapter 3 presents an overview of our survey and addresses some of the data issues encountered in our analysis. Chapter 4 provides a conceptual framework for the empirical analysis in the monograph and addresses various estimation issues. Chapter 5 presents the results of our test of theoretical validity. Similarly, Chapter 6 addresses the estimation issues and results of the tests for CV. Chapter 7 performs the tests of reliability. Finally, Chapter 8 examines the implications of our results from several perspectives ranging from litigation associated with oil spills to the development of new damage assessment regulations for oil spills.

CHAPTER 2

Experimental Design

2.1 Background

This chapter describes our experimental design for this study. To put the design into perspective, we begin by discussing the concept of accuracy and how it has been addressed in the CV literature. We also discuss the aspects of accuracy that are important for CV in general, along with the specific implications for using CV to measure nonuse damages. As part of this discussion, we consider whether the demands of a damage assessment impose more stringent requirements for accuracy than other CV applications.

The statistical concept of accuracy is the ability of a measure, such as CV measures of WTP, to conform to an accepted standard or true value. Cummings, Brookshire, and Schulze (1986) were among the first CV practitioners to consider the question of accuracy. They examined various comparisons of CV and indirect methods. This analysis leads Cummings, Brookshire, and Schulze to conclude that, under properly specified conditions, CV results are sufficiently reliable if the range of error is less than plus or minus 50 percent. They do not identify what applications they had in mind for which this range of uncertainty is sufficient. The authors also note that the conditions are most likely to be fulfilled for well-defined commodities such as use values. They are much less optimistic about the use of CV to measure nonuse values.

The landmark Cummings, Brookshire, and Schulze study, combined with the careful analysis provided by Mitchell and Carson (1989), motivates the experimental design for our evaluation of CV accuracy. In particular, we examine the available evidence on validity and reliability that are especially important to using CV in damage assessments. Our review cites past studies that seem to have the most relevance to damage assessment issues. We especially are concerned about how to interpret evidence from use studies to situations where nonuse values predominate.

Our review of accuracy is not intended to be all inclusive. Rather, its scope is limited to considering the implications of the results for damage assessment. This chapter indicates how our experimental design leads to hypotheses that will shed some light on a debate that at times has had too much heat and not enough substance.

2.2 Assessing the Validity of CV

The concept of validity provides standards for evaluating experimental designs and the empirical results based on those designs. Three forms of validity are relevant for assessing our experimental design: content, construct, and criterion validity.

Content Validity

The content of a CV survey is valid if the questions are phrased in a way that induces respondents to give unbiased statements about their WTP. Evaluations of content validity usually have involved qualitative assessments of the CV questionnaire. For example, Desvousges and Smith (1988) advocate using focus groups to evaluate questionnaire content, visual aids, and experimental design questions. Other techniques, including one-on-one, in-depth interviews and small-sample pretests, also provide information for such qualitative assessments (Smith and Desvousges, 1986b). Mitchell and Carson (1989) recommend such procedures as standard practice in a thorough CV study.

The goal of qualitative assessments of content validity is to determine whether the questionnaire meets the elusive criterion of "working." This criterion usually means that respondents understand the questions that are being asked and they are sensitive to the key elements in the contingent scenario. These qualitative assessments also try to determine whether or not respondents are being influenced by unintended cues. If such cues are present, the estimates of WTP may be biased and the survey will not fulfill the content validity requirement. (See Mitchell and Carson, 1989 and Carson, 1991 for further discussion.)

The evaluations of starting point bias by Boyle, Bishop, and Welsh (1988), payment vehicle bias by Milon (1989), and interviewer bias by Desvousges, Smith, and Fisher (1987) are quantitative assessments of content validity.[1] These studies attempt to determine if interviewers, payment vehicles, and starting points provide unintended cues to respondents. The results of these tests, combined with the emergence of other question formats, have led to the decline in popularity of bidding games, which were the standard in early CV studies. These methodological investigations often use more extreme treatment values than earlier studies to assess content validity.

More recently, several other studies have attempted to include broader tests of content validity in the experimental survey design. For example, Smith and Desvousges (1986a and 1987) compare differences in implied property rights in their experimental design to determine whether people respond to such differences. (This idea emerged from focus-group testing, which examined the workability of the questionnaires.) Their results show that people are very sensitive to the assignment of property rights, with payments for achieving a risk decrease being significantly larger than values for avoiding a comparable increase in risk. Carson (1991) and Carson, Hanemann, and Kopp (1992) vary the levels of visibility and health benefits included in different versions of their experimental design to determine if people respond only to visibility changes or if they also consider health benefits. They conclude that extensive pretesting produces a questionnaire that is able to isolate the visibility effects and that people respond appropriately to this dimension of the choice problem.

The CV literature appears to have devoted considerable attention to content validity. Many of the improvements in CV practice are aimed at improving the workability of questionnaires. The attempts to include workability tests into the experimental design of studies are crucial for using CV in NRDAs. It is essential to show that people respond to the questions as the analysts intend and that these responses correspond to conceptually correct measures of values.

[1] Content validity also is addressed by Cummings, Brookshire, and Schulze (1986).

Criterion Validity

Criterion validity requires that a CV survey be capable of producing "true" estimates of WTP for a specific commodity. It is evaluated by comparing WTP estimates to some objective, unestimated value. The most relevant criterion for CV studies is actual market prices. Researchers have used a wide variety of hypothetical-simulated market experiments to study criterion validity. These experiments are designed to test whether the values people state in a survey indicate what they really would pay if an actual market existed. These experiments generally are confined to use values, which is an important consideration in evaluating their relevance for measuring nonuse damages.[2]

Dickie, Fisher, and Gerking (1987) compare the estimates elicited in a CV survey to actual market prices for strawberries, a purely private good. Interviewers visited 144 randomly selected households in Laramie, Wyoming, and asked how many pints of strawberries a household would be willing to buy for a given price. Prices ranged between $.60 and $1.60 per pint. In comparing the estimated hypothetical and actual market-price demand curves for strawberries, they could not reject the null hypothesis that the two treatments are statistically identical. Their findings support the proposition that CV surveys achieve criterion validity (for private goods). If CV was not valid in this circumstance, there would be little hope for it in more difficult choice situations.

Brookshire and Coursey (1987) conducted an experiment involving the purchase of trees for a neighborhood park in Colorado. In this study, they compare values collected with a CV survey to values revealed through simulated auction markets using actual cash transactions. These auctions were conducted both in the field and in the laboratory. This comparison yields findings similar to the Dickie, Fisher, and Gerking study because the CV WTP responses do not diverge very far from the auction results. The Brookshire and Coursey study constitutes a somewhat broader test because their experiment involves a mixed public/private good. However, it also is a well-known good for which people have considerable choice experience, and there is some debate about the interpretation of the statistical results (see Cummings and Harrison, 1992).

[2] Seip and Strand (1990) and Duffield and Patterson (1992) have attempted market experiments for nonuse values. Both studies have experimental design problems that make their results difficult to interpret.

Heberlein and Bishop (1986) describe three different experiments to test the criterion validity of CV surveys. Their third experiment is the most conclusive. The study concerns buying and selling 1-day hunting permits for the Sandhill Wildlife Demonstration Area in Wisconsin. A key part of this experiment tests "intended behavior" WTP estimates. A sample of hunters was chosen from the applicants who did not receive a Sandhill permit in the lottery. Half the sample was offered a contract that would allow them to buy a permit for the amount designated in the contract. The other half received identical offers, except these offers were hypothetical. Heberlein and Bishop find no statistically significant difference between the hypothetical-market WTP and the actual or simulated-market WTP.

Heberlein and Bishop conclude that while hypothetical, markets are capable of valuing well-defined goods with considerable accuracy, their results are only weak evidence that hypothetical markets can achieve criterion validity for such less well-defined goods as the value of many nonuse services in NRDAs. Criterion validity is an essential element for appraising accuracy, especially in NRDAs in which someone must write a check at the end of the process. Establishing whether people actually would make the hypothetical payments that are elicited in the CV questionnaire is not an unreasonable requirement. As we discuss in Chapter 8, criterion validity clearly is an area where further research is needed.

Construct Validity

According to Carmines and Zeller (1979), construct validity is critical for assessing measures of abstract theoretical concepts. Ideally, construct validation requires a pattern of consistent findings involving different designs across a number of studies. Construct validity can be divided further into two categories: theoretical validity and convergent validity.

Theoretical Validity

Theoretical validity concerns the extent to which a particular measure is consistent with theoretically derived hypotheses. Carmines and Zeller (1979) outline three steps for assessing theoretical validity:

- specifying the theoretical relationships,
- examining the empirical relationships, and

- interpreting the empirical evidence in terms of how it clarifies the construct validity of the particular measure.

This approach emphasizes deriving testable hypotheses from a well-developed theoretical basis. Freeman (1990) also develops criteria that relate to theoretical validity:

- The framework should be consistent with the standard economic theory of individual preferences and measurement of welfare changes.
- The theoretical framework and the measurement techniques derived from it should be capable, at least in principle, of reproducing what would be revealed as values by an actual market for the resource, if such a market existed.
- The theory and measurement techniques should be resource specific, rather than reflecting general attitudes.

According to Freeman, a CV survey must fulfill each of these three requirements to be capable of producing valid, and ultimately accurate, estimates of WTP.

Boyle and Bishop (1988) test the theoretical validity of CV. Their experiment assesses whether people hold nonuse values for the bald eagle. In two independent surveys, they asked two versions of a WTP question about preventing the extinction of the bald eagle in Wisconsin. They asked one-half of the sample to estimate their total value for the bald eagle without any conditions being placed on this value. The other half of the sample was asked to state their WTP if the bird habitat was in remote parts of the state where viewing is not possible. Both valuation questions apply to current users and nonusers. The CV question for the second sample, however, factors out any possible use component of total value since it stipulates that viewing will not be possible. As expected from theory, Boyle and Bishop find that the WTP values elicited from the conditional valuation question are less than the WTP values elicited from the unconditional valuation question and that the conditional value is greater than $0.

Smith and Desvousges (1987) also examine theoretical reliability in their study of hazardous waste risks. Using a detailed experimental design, and a relatively large sample (609 respondents), they formulate hypotheses based on expected utility theory about the relationships between option prices and changes in risk from exposure to hazardous wastes. They find that people's responses are not consistent with the conventional predictions of expected utility theory for a compound lottery. However, this finding could have two interpretations: either the CV survey was not valid or the conventional expected utility theory is not consistent with people's actual behavior under uncertainty. Given the substantial body of literature that has raised questions about the predictive ability of expected utility theory, this is at least an open issue (see Hogarth and Reder, 1986).

Mitchell and Carson (1989) suggest another approach to assessing theoretical validity. They suggest regressing the WTP estimates from a CV survey on a group of independent variables believed to be determinants of WTP. They argue that such variables as income, current levels of consumption, and existing tastes and preferences should have a theoretically consistent relationship to WTP. For example, economic theory requires that people be willing to pay more for greater air quality improvements than for lesser ones. The size and sign of the estimated coefficients should be examined to determine whether the observed relationships are consistent with economic theory. If the estimated relationships are not observed as economic theory dictates, the theoretical validity of the measure may be questionable. The Desvousges, Smith, and Fisher (1987) regression analysis that examines the role of income in option price estimates is an example of the Mitchell and Carson view of theoretical validity.[3]

Theoretical validity is a basic prerequisite for CV to be useful in damage assessments. This criterion must be met in order for the estimated values to correspond to the conceptually correct measures of damages upon which the NRDA regulations are based. As we discuss later in this chapter, the experimental design developed for this study examines theoretical validity in three independent tests.

Convergent Validity

Convergent validity requires WTP estimates for given levels of a commodity to be the same for different elicitation frameworks. The majority of studies that have assessed construct validity have concentrated on convergent validity. Convergent validity has been tested by comparing CV estimates to estimates produced by a

[3] Mitchell and Carson (1989) cite many more studies that somehow test theoretical validity.

so-called indirect method such as travel cost or hedonic pricing.[4] It is important, however, that these comparisons be carefully performed. For example, the travel cost model produces an ex post consumer-surplus estimate of a welfare change, so the relevant CV comparison must provide the same welfare measure. It also is important to recognize that neither the travel-cost or hedonic-pricing measure is the "true" measure. Both are approximations of the true welfare change.

One example of a study testing convergent validity is Smith, Desvousges, and Fisher (1986). They conducted a study that compares alternative approaches for estimating the benefits of water quality improvements. Their study is based on a case study of the Monongahela River and focuses specifically on the travel cost and CV approaches for estimating WTP. They find the following:

- CV estimates of water quality improvements overstate WTP, although their results do not indicate a statistically significant difference between the two sets of estimates.
- The travel cost model overstates WTP for the loss of an area, and the estimates are not comparable to the CV estimates.
- The comparative performance of CV relative to the travel cost method is sensitive to differences in question format used in the CV survey.
- The null hypothesis that there is no association between methods is clearly rejected at high levels of significance.

Smith, Desvousges, and Fisher show a level of correspondence between CV and travel-cost estimates on the order of 50 to 60 percent. This result is consistent with the Cummings, Brookshire, and Schulze (1986) review of comparison studies. Finally, the studies that have examined convergent validity have all involved measures of use values. As noted for the tests of criterion validity, the implications of comparisons of use value estimates for nonuse values are unclear.

[4] Convergent validity studies comparing indirect valuation methods with CV include Knetsch and Davis (1966); Bishop and Heberlein (1979); Thayer (1981); Brookshire et al. (1982); Bishop, Heberlein, and Kealy (1983); Desvousges, Smith, and McGivney (1983); Cummings, Brookshire, and Schulze (1986); Seller, Stoll, and Chavas (1985); Heberlein and Bishop (1986); and Smith, Desvousges, and Fisher (1986).

A second aspect of convergent validity is whether the measures are consistent across different measurement techniques. For CV, one relevant test of convergent validity is whether the results are sensitive to changes in the question format. Smith, Desvousges, and Fisher (1986) compare open-ended format, payment card, and bidding games. Although they find that CV is sensitive to differences in question format, the implications for our current research are somewhat harder to draw because bidding games have fallen out of favor in current practice and payment cards are used infrequently. Unfortunately, there have been very few independent tests of the effect of question format for open-ended and dichotomous-choice formats, which are widely used in the current literature.[5] With the rapid increase in the use of dichotomous-choice questions, it is odd that this popularity has been achieved with relatively little evidence on its validity. Our study remedies this deficiency by using independent comparisons of open-ended and dichotomous-choice formats for values that are predominantly nonuse.

2.3 Assessing Reliability

Reliability is the second dimension of accuracy. Assessing reliability requires distinguishing between “true” variation and measurement error. Economic theory acknowledges that there will be some inherent or true variation in different people’s value for the same commodity. This variation is a result of differences in the determinants of WTP or in some underlying stochastic process. For example, income, quantity consumed, and existing tastes and preferences all influence demand and, therefore, WTP. True variation in WTP resulting from differences in these determinants will not affect the reliability of a WTP value determined by a proper measurement technique.

Measurement error, however, may substantially affect the reliability of a given technique. Measurement error refers to a difference between true WTP and estimated WTP that is attributable to deficiencies either in the survey questions or estimation procedures that are used to obtain the WTP estimates. Deficient survey questions can lead to measurement problems ranging from unintended or biased

[5] Johnson, Bregenzer, and Shelby (1990), Seller, Stoll, and Chavas (1985), and Kristöm (1988) are three exceptions. Other studies have asked both open-ended and dichotomous-choice format questions of the same respondents.

WTP responses to incorrect reporting of independent variables used in the analysis. For purposes of this monograph, we have chosen to address the issues related to measurement error in the survey questions as being associated with validity, and not reliability. The rationale for this choice is mainly to simplify the exposition. For reliability issues, we will concentrate on the potential measurement errors that relate to the statistical estimation procedures. As this discussion suggests, the lines between reliability and validity are hazy.

One measure of reliability is the R^2 value of a regression equation. Mitchell and Carson (1989) state that one must question the reliability of a CV study "which fails to show an R^2 of at least 0.15, using only a few key variables." Although 0.15 is an arbitrary standard, a high R^2 indicates that the unexplained portion of the variability in stated WTP is relatively small, which suggests less risk of measurement error. For theoretical reasons, the independent variables of the WTP regression equation need to be chosen carefully. In addition to affecting the validity of estimates, the choice of independent variables also affects the reliability of the estimate. However, R^2 is not as useful for assessing reliability with current techniques for analyzing CV data. Most studies no longer use classical regression analyses because of the censoring of the stated WTP at zero or because some type of dichotomous-choice model is estimated. The conventional R^2 statistic cannot be calculated for these models. Although other goodness-of-fit measures are available for judging the predictive ability of these models, their interpretation is more difficult than for classical ordinary least squares.[6] (See Desvousges, Smith, and Rink [1988] for a discussion of this issue.)

The experimental design literature suggests that measurement error can be measured by testing and retesting the same sample using an identical questionnaire. Kealy, Dovidio, and Rockel (1988) performed an experiment using a purely private good, Cadbury chocolate candy bars. The experiment is designed to

- assess the stability of CV values in hypothetical situations across time (reliability),
- test different methods of valuing the same good (dichotomous-choice vs. open-ended), and
- examine the discrepancy between CV values and actual purchase behavior (hypothetical simulated-market).

[6] Alternative goodness-of-fit measures include McFadden R^2, pseudo-R^2, and log-likelihood ratio.

In this study, 2 weeks elapsed between the administration of the CV questions. The test-retest results indicate that both dichotomous-choice and open-ended question formats are reliable. Critics would argue, however, that the results of this study are limited for two reasons. First, most people have purchase experience with chocolate candy bars, which makes them relatively easy to value. Second, the time lapse of 2 weeks may not be sufficient. For a study to properly test reliability, respondents should not be able to recall their initial estimates (Reiling et al., 1990).

Loomis (1989) performed a test-retest study in order to assess the reliability of WTP estimates for preservation values. His study compares the consistency of WTP values for the preservation of Mono Lake in California. The test-retest procedure involved mailing an identical questionnaire to the same people in April 1986 and then again in January 1987. Using only the questionnaires of respondents answering both surveys, Loomis shows that respondents consistently gave the same WTP estimates (the correlation was about 0.70), and their WTP values did not change significantly over time. Although the consistency of these responses indicates a degree of reliability, this result does not imply that the responses also are valid.

This same study also tests the comparative reliability of open-ended and dichotomous-choice CV formats. The surveys ask a series of open-ended WTP questions to determine the respondents' maximum WTP as well as their marginal WTP. The surveys also ask a series of dichotomous-choice questions to determine whether respondents are willing to pay specified amounts. Loomis' study reveals no significant difference in the test-retest correlations between the open-ended and dichotomous-choice question formats. It is important to note, however, that the answers given to two sets of questions in the same survey are not independent. The lack of independent samples severely limits the generalizability of the comparative reliability tests.

Finally, reliability can be examined by testing the sensitivity of the WTP responses to variations in the statistical analyses. For example, Boyle (1990) and Bowker and Stoll (1988) test the sensitivity of dichotomous-choice estimates to differences in functional form of the estimation model. Similarly, it is possible to examine changes in model specification. The evidence from such investigations indicates the reliability of the CV estimates. Reliability requires that WTP estimates be robust for competing models that are not clearly superior to each other on either statistical or theoretical grounds.

Reliability tests can also be included in the basic survey design. The studies that have used repeated measurements of WTP over time are a good example. Reliability tests can also be performed after the data are collected to address the robustness of the statistical results to choices made by the analyst. The key reliability issue for the latter type of tests is whether different researchers using the same data would get significantly different estimates of WTP (see Desvousges and Dunford, 1991; and Carson, Hanemann, and Kopp, 1991).

Reliability concerns the sensitivity of an estimate to the choice of estimation method. Although reliability is especially important for NRDAs, it cannot be considered independently of validity. However, it does not establish whether that answer is a valid estimate of WTP. As the current NRDA regulations now stand, no guidance is offered on what types of tests are needed to provide both valid and reliable estimates of WTP.

2.4 Experimental Design

The experimental design is an essential component of any test of CV accuracy. Figure 2-1 shows the two-part design of this study. In the first part of the design, we use three versions of the migratory-waterfowl questionnaire. These versions, which are identical in all other respects, indicate one of three levels of migratory waterfowl protection: 2,000, 20,000, or 200,000 deaths prevented annually. Simply put, this experiment tests whether the WTP estimates increase as the number of protected birds increases by two orders of magnitude. All three versions use an open-ended format for the WTP question, and the respondents are asked for total value. From the standpoint of the experimental design, the only difference in the three versions of the questionnaire is the number of bird deaths described in the scenario. The questionnaires are identical in all other respects, yielding a 3 × 1 experimental design.

The second part of the study incorporates a 2 × 2 experimental design for the oil-spill experiment, with two types of spill response (small spills only and all spills) and two WTP question formats (open-ended and dichotomous-choice). Again, the WTP questions ask for total value. The all-spills protection includes the small spills and adds the incremental response for larger spills. This type of embedded design differs from that used by Kahneman and Knetsch (1992), in which the relationship between design points is difficult to interpret. Two different questionnaire formats are used in the oil-spill experiment for both the all-spills and small-spills design. Using independent samples for the two question formats, which has been lacking in many previous evaluations, is essential for the hypothesis testing.

As shown in Figure 2-1, the goal was 400 interviews in each open-ended cell and 390 interviews in each dichotomous-choice cell. The dichotomous-choice versions randomly assigned respondents one of six bids: $10, $25, $50, $100, $250, and $1,000.[7] The sample sizes were chosen to balance budgetary considerations with the need to have sufficient power for being able to detect differences. These sample sizes are at least comparable, and in many cases much larger, than those used in previous CV experiments (see Mitchell and Carson, 1989).

Figure 2-1. Experimental Design with Target Number of Interviews

Migratory-Waterfowl Protection

	2,000 birds	20,000 birds	200,000 birds
Open-Ended Question Format	400	400	400

Oil-Spill Response

	Small Spills	All Spills
Open-Ended Question Format	400	400
Dichotomous-Choice Question Format	390	390

Our experiments are designed specifically to test three simple hypotheses regarding the accuracy of CV for values that have a substantial nonuse component:

- *Hypothesis 1*: WTP estimates increase for higher levels of natural resource services. This hypothesis is a test of theoretical validity.

[7] See Chapter 7 for discussions of the procedure used to determine the bid structure and the sensitivity of results to changes in the bid structure.

- *Hypothesis 2*: The open-ended and dichotomous-choice CV formats yield comparable WTP estimates for the same natural resource services. This hypothesis is a test of convergent validity.
- *Hypothesis 3*: The dichotomous-choice WTP estimates are not sensitive to reasonable choices in the analysis of the data. This hypothesis is a test of reliability.

Hypothesis 1 requires that the estimates produced by a CV survey be consistent with economic theory. One test of theoretical validity, as discussed earlier, is to elicit WTP values for different levels of the same commodity. In our first experiment, we elicit stated WTP values for three levels of waterfowl protection: 2,000 birds, 20,000 birds, and 200,000 birds. Respondents are asked to state their WTP to prevent one of these numbers of birds from dying each year in waste-oil holding ponds. Theoretical validity requires that the estimated WTP values increase as the number of birds affected increases.

In our second experiment, WTP values are elicited for protecting the environment from either small oil spills only or all oil spills. Comparing the WTP values for the different levels of spill protection reveals whether people are willing to pay more to prevent all oil spills than to prevent small oil spills alone. It is important to emphasize that the small-spill protection is completely nested in the all-spills protection to make the comparison a straightforward test of theoretical validity. Additionally, the oil-spill response centers provide us with an opportunity to see how well CV performs with a contingent commodity that is immediately relevant to NRDA. Oil spills, unlike birds in waste-oil ponds, are an emotional issue with which respondents are familiar before they take the survey. The baseline information that they have received from various sources also may be of questionable accuracy. Thus, respondents may feel that our survey is an opportunity to make a statement about an important public policy issue.

The Hypothesis 2 test evaluates convergent validity by comparing open-ended responses with dichotomous-choice responses for the same commodity—oil-spill protection. Convergent validity requires that the WTP estimates for given levels of nonuse commodities be comparable, regardless of how the valuation question is asked. Given the dramatic increase in the use of dichotomous-choice questions in

CV surveys, the test of Hypothesis 2 is an important contribution to understanding the accuracy of CV estimates.

The Hypothesis 3 test assesses the reliability in the dichotomous-choice WTP estimates with respect to estimation issues such as functional form and bid structure. According to theory, these issues should not unduly influence the variability of the estimates. Each of these factors is at the discretion of the analyst, so an evaluation of their effects on WTP is especially important for addressing reliability.

Finally, these hypotheses are fundamental to the use of CV for measuring losses such as those associated with oil spills. Each hypothesis tests one aspect of the accuracy of CV estimates, including theoretical and convergent validity and reliability. The commodities are defined such that nonuse values are likely to be a significant share of the total value. If CV is unable to yield accurate WTP estimates for oil spills of different magnitudes or differing levels of waterfowl protection, then it has limited relevance for damage assessment.

CHAPTER 3

Survey Overview and Data Issues

3.1 Background

The quality of CV data depends on the quality of the survey instrument. The survey must be well-designed and thoroughly tested. Additionally, an administration mode must be selected that balances budgetary constraints with the need for data that are reliable. After the data are collected, some consistency checks and "data cleaning" must be performed to ensure the integrity of the data set. This data cleaning process involves judgments on the part of the analysts as to the quality of the data. Some of these judgments have evolved into common practice procedures that are generally used in CV procedures, while others are more controversial and require more discussion. Noticeably lacking in this area is a well-developed theory that would provide a consistent rationale for making decisions about what responses should be included or excluded from the data set.

This chapter presents an overview of the surveys administered in our nonuse experiments, including some summary information on the WTP responses from the surveys. Also, this section discusses protestors and outlier responses. Appendices B, C, and D contain the migratory-waterfowl and two oil-spills questionnaires.

3.2 Questionnaire Design

This section describes the organization and content of the migratory-waterfowl and oil-spills questionnaires. The goals of both questionnaires are:

- to elicit unbiased WTP responses, and
- to obtain information on respondent characteristics that are likely to influence their WTP responses for the contingent commodity.

A well-designed and thoroughly tested questionnaire is a prerequisite for any survey. This is particularly true for CV studies where respondents are asked to answer difficult questions regarding tradeoffs between money and the quantity or quality of a public good. As noted in Chapter 2, thorough pretesting of CV surveys enhances the content validity of the entire survey. CV proponents have criticized studies, such as Kahneman and Knetsch (1992) and Seip and Strand (1990), for weaknesses in both their questionnaires and their experimental design (see Smith, 1992). This chapter describes the intensive efforts devoted to designing our questionnaires.

Migratory-Waterfowl Questionnaire

The migratory-waterfowl questionnaire has five sections:

- introduction and background information on migratory waterfowl;
- information on waste-oil holding ponds, deaths of migratory waterfowl in these ponds, and proposed (hypothetical) federal regulations requiring owners to cover these ponds with a heavy wire net;
- WTP question and follow-up probes;
- opinion/attitude questions; and
- socioeconomic questions.

Appendix B contains a complete copy of the migratory-waterfowl questionnaire. For the actual survey administration, the questionnaires were printed in a booklet format using four, legal-sized pieces of colored paper folded in half and stapled in the fold.

The first section of the migratory-waterfowl questionnaire contains introductory and background information on migratory waterfowl. It includes the following:

- a definition of migratory waterfowl,
- a question on respondents' recent exposure to information on migratory-waterfowl issues,
- a question on the importance of migratory waterfowl to the respondent and his/her reasons for wanting to protect migratory waterfowl (if applicable),
- a map and accompanying information on the four main migratory-waterfowl flyways in the United States (emphasizing the Central Flyway), and
- a question on respondents' knowledge of various threats to migratory waterfowl in the Central Flyway (including uncovered waste-oil holding ponds).

This section of the questionnaire has three goals. The first is to convey basic information regarding migratory waterfowl, the various flyways (with emphasis on the Central Flyway), and potential threats to migratory-waterfowl populations. The second goal is to stimulate respondents to consider their personal knowledge of migratory waterfowl, and to reflect on their preferences regarding the preservation of migratory waterfowl. Finally, we want respondents to provide us with information on their knowledge and preferences that can be used in analyzing responses to the CV question. These three goals are accomplished concurrently by providing respondents with short information cues and then asking them to respond to questions. Both the information and questions were designed to inform respondents about the object of valuation and to induce them to consider their preferences.

The second section of the questionnaire provides information on deaths of migratory waterfowl in waste-oil holding ponds in the southwestern United States. The number of deaths is presented in absolute and relative numbers to help respondents fully understand the magnitude of migratory-waterfowl deaths in the waste-oil holding ponds in the Central Flyway each year. The practice of giving both types of information is consistent with the findings from focus groups conducted for a previous study (see Desvousges and Smith, 1988). In those focus groups, respondents often calculated the percentages when given only absolute information or information expressed as a fraction.

As indicated earlier, we administered three versions of the survey to test whether respondents' evaluations are sensitive to the number of annual migratory-waterfowl deaths. The three versions are identical except for the number of migratory waterfowl that die in the holding ponds each year. Specifically, the three versions vary as follows:

- In 1989, for example, about **2,000** migratory waterfowl died in these holding ponds. This was much less than 1 percent of the 8.5 million migratory waterfowl in the Central Flyway.
- In 1989, for example, about **20,000** migratory waterfowl died in these holding ponds. This was less than 1 percent of the 8.5 million migratory waterfowl in the Central Flyway.
- In 1989, for example, about **200,000** migratory waterfowl died in these holding ponds. This was about 2 percent of the 8.5 million migratory waterfowl in the Central Flyway.

Respondents were randomly assigned a survey containing one of these three treatments.

The second section of the questionnaire also provides information on the commodity to be evaluated and proposes (hypothetical) federal regulations requiring owners to cover their waste-oil holding ponds. Specifically, respondents are told that these regulations would require covering these ponds with a heavy, small-mesh wire netting that would prevent the migratory waterfowl from coming in contact with the oily wastes.[1] The regulations would have the USFWS monitor compliance with the regulation and cover abandoned ponds.

After describing the commodity, the questionnaire explains that oil companies would ultimately pass on the costs of wire-net covers to consumers in the form of higher prices for petroleum products, which would increase the price of most other things that they buy. This explanation is followed by a statement in bold letters: **It is important to know how much protecting these migratory waterfowl is worth to you**. We include this statement to help the respondent focus on the value they place

[1] In our pretests we showed participants drawings of the proposed wire-net covers, but the drawings confused them. Consequently, we decided to eliminate them in the final questionnaires.

on preventing migratory-waterfowl deaths. We do not want respondents to try to estimate the expected cost of the regulation or to react to the acceptability of higher oil prices.[2]

Respondents are then told to think about their current household income, current household expenses, and other possible uses for household income. Following these statements, we ask an open-ended, CV question, phrased as follows:

> Keeping these factors in mind, what is the **most** that your household would agree to pay **each year** in higher prices for wire-net covers to prevent about 2,000 (or 20,000 or 200,000 depending on the version) migratory waterfowl from dying each year in waste-oil holding ponds in the Central Flyway?

The question does not ask how much people would be "willing to pay." Our pretests determined that this particular phrase confuses respondents. Instead, we ask how much they "would agree to pay."

As noted, the only part of the question that differs across the three versions of the survey is the number of birds protected. Therefore, any changes in the values across versions should be attributable to the change in the number of birds.

The key features of this question include the following:

- a reminder of the annual or yearly nature of the payments using bold-faced type;
- a reminder of the need for a maximum amount also in bold;
- a reminder of the number of birds deaths that would be prevented and their location; and
- a clear definition of the policy option that would prevent the bird deaths—the covers for the waste-oil holding ponds.

These features ensure that this question meets the specifications that Cummings, Brookshire, and Schulze (1986) propose for valuation questions, as well as those suggested by Fischoff and Furby (1988).

Exploratory probe questions follow the CV question to determine respondents' most important reason for their answer and to help identify invalid WTP responses. Separate sets of probes are provided, depending on whether the respondent's answer

[2] Our pretests indicated that respondents often try to assess the cost of a policy, not the value.

to the open-ended question is a positive, nonzero number or something else. We used these responses to identify invalid responses, as explained later in this section.

The fourth section of the migratory-waterfowl questionnaire contains two sets of opinion statements. In one set, respondents are asked to indicate the extent of their agreement or disagreement with the general statements regarding their environmental beliefs. These statements address whether:

- people have the right to change the environment to meet their needs,
- people have to make choices between a clean environment and a strong economy, and
- the environment should be preserved at all costs.

The second set consists of two statements related to respondents' beliefs about the seriousness of migratory-waterfowl deaths and whether the wire-net covers would be effective. These questions are specifically designed to explore respondents' opinions that may affect their valuation of migratory waterfowl.

The final section contains the typical sociodemographic questions about age, education, sex, race, income, and number of people in the household. In addition, this section requests information on respondents' recreational activities involving wildlife, in general, and migratory waterfowl, in particular, in the Central Flyway region and elsewhere in the United States. These questions identify past and present users of the migratory waterfowl, as well as users of other wildlife. The final section also asks respondents if they are members of several selected environmental or conservation organizations. Membership might be an indication of a person's environmental awareness, interest, and revealed WTP for wildlife and environmental quality.

Oil-Spills Questionnaire

The oil-spills questionnaire also consists of five parts:

- opinions and attitudes about the importance of various public policy issues, in general, and environmental issues, in particular;
- information on the CV commodity (i.e., oil-spill response centers that would limit the environmental effects of oil spills);
- WTP question and follow-up probes;

- opinion/attitude questions; and
- socioeconomic questions.

Appendix C contains a complete copy of the oil-spills questionnaire using the open-ended format, and Appendix D contains the oil-spills questionnaire using the dichotomous-choice format. Both questionnaires were administered in a booklet format using four legal-sized pieces of colored paper folded in half and stapled in the fold.

The first section begins by asking respondents to rank the seriousness of various social and economic problems in the United States, including environmental pollution. The next question asks respondents for a low, medium, or high priority for government funding to address several pollution sources, including oil spills from tankers. The third and last question in this section focuses on respondents' recent exposure to information on oil spills. This section serves to start the respondent thinking about public policy issues in general, and oil spills in particular. In addition, it allows us to categorize respondents based on their general environmental opinions and their level of knowledge about oil spills. Because these answers precede the commodity description, we are able to obtain some indication of the respondents' baseline opinions.

The second section provides information on the Oil Pollution Act (OPA) of 1990, the causes of oil spills, and the injuries that typically result from oil spills. Two knowledge/opinion questions about the OPA separate some of the information paragraphs in order to maintain the respondents' interest. In the small-spills version of the questionnaire, respondents are told that the Federal government is considering regulations requiring oil-spill response centers at every U.S. port that handles oil. These local response centers are then described, and respondents are told that these response centers will prevent 90 percent of the environmental damage from oil spills of less than 50,000 gallons, but they will not be effective in preventing environmental damage from larger oil spills. The all-spills version contains the same description of local response centers for small oil spills, as well as a description of regional response centers that would prevent 75 percent of the environmental damage from spills involving more than 50,000 gallons of oil. Therefore, the small-spills scenario is completely embedded in the all-spills scenario.

While developing the questionnaire, it was necessary to decide whether to describe how the response centers would affect the risks of an oil spill. Given our past experience with the difficulties of communicating risk information in a CV survey and the problems people have in answering WTP questions when risks are involved, we decided to describe the effectiveness of the centers in percentage terms (Smith and Desvousges, 1987). That is, the small-spills scenario would prevent 90 percent of the environmental damage from oil spills less than 50,000 gallons. The all-spills scenario would prevent 75 percent of the environmental damage from oil spills greater than 50,000 gallons while also preventing 90 percent of the environmental damage from spills less than 50,000 gallons. In pretests we found that asserting that these centers would completely eliminate the environmental effects of oil spills was not credible to the respondents. We use a national scope for the oil-spill response centers because we found it to be a more realistic way to portray the types of issues that arise in spills rather than focusing on only one oil spill. We evaluated this dimension of the questionnaire, along with the amount of information about the baseline regulations already in place, in a focus group and several one-on-one interviews.

The third section of the oil-spills questionnaire contains the WTP question and the follow-up probes. The WTP question tells respondents that oil companies would pass on the cost of oil-spill response centers to consumers in the form of higher prices for petroleum products, which would increase the prices of most other goods and services. (This is the same payment vehicle used in the migratory-waterfowl questionnaires.) This explanation is followed by a statement in bold letters that respondents' values for reducing the environmental effects of oil spills are important, which is indented to help respondents focus on the value they place on reducing environmental effects from oil spills, not the expected cost. The respondents are then told to think about their current household income, current household expenses, and other possible uses for their household income, and asked the valuation question. Table 3-1 shows the wording for each of the valuation questions.

In all versions of the oil-spills questionnaire, the WTP question is followed by probes to determine respondents' most important reason for their answer. As mentioned above, these probes are intended to identify invalid WTP responses. As in the migratory-waterfowl surveys, for the open-ended versions, respondents answer different sets of probes according to whether their response to the valuation question is a positive amount or something else. Similarly, the dichotomous-choice respondents answer different probes, depending on their response to the WTP question. Additionally, the respondents who answer No to the dollar amount in the dichotomous-choice questionnaires are also asked if their household would agree to pay **any amount** to reduce the effects of oil spills. This screening question attempts to identify those dichotomous-choice respondents who have a value of zero.

Table 3-1. Wording of Oil-Spill Contingent Valuation Questions

It is important to know how much reducing the effects of oil spills is worth to you.	
	Please think about: • Your current household income • Your current household expenses • Other possible uses for your household income
Open-Ended, Small Oil Spills	Keeping these factors in mind, what is the **most** that your household would agree to pay **each year** in higher prices for local response centers to reduce the effects of oil spills *less than* **50,000 gallons**?
Open-Ended, All Oil Spills	Keeping these factors in mind, what is the **most** that your household would agree to pay **each year** in higher prices for local and regional response centers to reduce the effects of **all** oil spills?
Dichotomous-Choice, Small Oil Spills	Keeping these factors in mind, would your household agree to pay **$XX**[a] more **each year** in higher prices for local response centers to reduce the effects of oil spills *less than* **50,000 gallons**?
Dichotomous-Choice, All Oil Spills	Keeping these factors in mind, would your household agree to pay **$XX**[a] more **each year** in higher prices for local and regional response centers to reduce the effects of **all** oil spills?

[a] The dichotomous-choice surveys randomly assigned six bids: $10, $25, $50, $100, $250, and $1,000.

The fourth section of the oil-spills questionnaire contains three sets of opinion questions. The first set of opinion questions asks respondents to rate the seriousness of four recent US oil spills. In the second set of opinion questions, respondents are asked to indicate the extent of their agreement or disagreement with three statements about tradeoffs between the environment and economic activities. (These same statements are also in the migratory-waterfowl questionnaires.) Finally, respondents are asked to indicate the extent of their agreement or disagreement with three statements about oil spills.

The fifth section of the oil-spills questionnaire obtains socioeconomic information on respondents, such as age, education, sex, race, income, and number of people in the household. This section also asks respondents if they visit the beach or other coastal areas for recreation in a typical year, in order to identify respondents who may include some prospective use values in their WTP response. The last question in this section asks about respondents' membership in several environmental or conservation organizations.

3.3 Questionnaire Development

To develop the survey instruments, we engaged in an extensive development process over a 3-month period. As shown in Table 3-2, we used a range of techniques in developing the questionnaires, including focus groups, several one-on-one pretests, and two mall pretests on each questionnaire. The mall pretests served as field tests of

Table 3-2. Summary Information on Questionnaire Development Activities

		Number of Participants	
Questionnaire Development Activity	Date	Migratory Waterfowl	Oil Spills
Focus Group	3-11-91	10	10
One-on-one Pretest	3-14-91	6	6
One-on-one Pretest	3-20-91	5	6
One-on-one Pretest	3-27-91	—	9
Mall Pretest (Raleigh, NC)	4-9-91	146	208
Mall Pretest (Atlanta, GA)	5-8-91	32	69

the survey instrument because we ultimately drew our sample from shoppers at two malls in Atlanta, Georgia. In total, we administered various drafts of the migratory-waterfowl and oil-spill questionnaires during this period to 199 and 308 people, respectively. This pretesting is more extensive than in past studies (see Smith and Desvousges, 1986b; and Smith, Desvousges, and Freeman, 1985). The number of pretest surveys administered approaches the sample size of the Monongahela study (see Desvousges, Smith, and McGivney, 1983).

We found this questionnaire development process to be very useful and informative. The focus groups, which were lead by project staff, allowed us to explore preliminary ideas for the questionnaires, and to learn how potential respondents viewed the commodities we would be asking them to evaluate. We discovered that group members expressed some confusion about the details of the proposed policies, as well as some suspicion about their effectiveness. The focus groups also clearly demonstrated that people have some preconceived notions about oil spills that are strongly held. Many of these notions are not based on factual information. By asking the group members to discuss how they determined their WTP amount, we found that many people were estimating the expected cost of the programs, rather than considering their own marginal values. Instead of telling us how much the commodity was worth to them, they were telling us how much they thought it would cost. As a result, we added some language to the WTP question to focus respondents' attention on the value of the commodity.

The one-on-one interviews, which also were run by project staff, provided the opportunity to probe individual's responses in detail without worrying about group dynamics. Respondents completed the questionnaire, and then they went through the survey with the interviewer, question by question, to get specific reactions. Over the course of several sessions, it became clear that some portions of the survey were working and some portions still needed adjustment. For example, several one-on-one migratory-waterfowl participants expressed concern that children also might be falling into the waste-oil holding ponds. They were focusing on that aspect, not on the deaths of migratory waterfowl, when answering the CV question. In the next version of the questionnaire, we included language to assure respondents that these ponds are located in "remote and sparsely populated areas." As noted earlier, we also determined that drawings of the proposed wire-net covers confused participants rather than helped them to understand how the covers would work. So, we did not use the drawings in the final migratory-waterfowl questionnaires.

In an early version of the small-spills questionnaire, spills of less than 50,000 gallons were categorized as "small." Many pretest participants were bothered by the term "small." They felt that it implied that these spills were unimportant, not worthy of attention. Consequently, we decided to remove the word "small." By the end of the one-on-one process in late March, we were satisfied that the commodity descriptions were clear and that respondents were able to complete the survey.

The mall pretests allowed us to test the operational details of the survey administration. For the first mall pretest, we administered 354 surveys in Crabtree Valley Mall in Raleigh, North Carolina, using recruiters from L&E Research, a market research firm in Raleigh. This round served three functions. First, we were able to make sure that the survey "worked" in the mall setting. Secondly, we collected enough data to enable us to do some preliminary analysis. The data were keyed into machine-readable form, and simple descriptive statistics were developed. These statistics allowed us to make qualitative decisions regarding how well individual questions might be working. This evaluation included examining the item non-response rate for key questions and the distributions of answers to questions to assess the spread of responses among the alternative answers. This analysis, while very preliminary, was quite informative. Finally, we used the oil-spill data, which included responses to the open-ended valuation question, to select our dichotomous-choice offer bids. Our final mall pretest was a production "dry-run" to simulate the final administration. Project staff trained recruiters and supervised the pretest in Atlanta.

Overall, we performed more pretesting on the oil-spill questionnaires than on the migratory-waterfowl questionnaires. The additional pretesting was necessary to refine the complex commodity description and to cope with the confounding effects of incorrect preconceived notions held by many respondents. It also allowed refinement of the dichotomous-choice valuation question.

3.4 Questionnaire Administration

Following training sessions conducted by RTI, Jackson Associates of Atlanta, Georgia, recruited survey participants in two large shopping malls in the Atlanta metropolitan area. Two malls were used to expedite the data collection. We chose

Atlanta because it is a large city outside the Central Flyway and is not located on a coast. Given the nonuse focus of the experiments, we wanted a low probability of recruiting users of migratory waterfowl in the Central Flyway. Although most of our respondents were users of coastal resources for recreation, the national policy context allows for a substantial nonuse component because it would protect all beaches, not just those used by the respondent.

Recruited participants were people over the age of 19 and were not full-time students. We established these screening protocols based on our pretesting experience. We decided to select adults who were in a position to make household decisions. Full-time college students also were eliminated because they often consider their parents' income when making expenditures, and their WTP responses may not be true reflections of their own preferences and budget constraints. We did not give any participants an incentive payment.

Table 3-3 shows that we had about 400 completed questionnaires for each of the three migratory-waterfowl versions and about 400 completed questionnaires for each of the four oil-spills versions. The distribution of questionnaires and dates of administration for the two malls were:

- Southlake Mall—1,923 completed questionnaires between May 17 and June 5, 1991
- Lakeshore Mall—889 completed questionnaires between May 22 and June 5, 1991.

Table 3-3. Experimental Design and Number of Completed Questionnaires for Each Design Alternative

	2,000 Birds	20,000 Birds	200,000 Birds	Total
Migratory Waterfowl:				
Open-ended format	398	408	399	1,205
		Small Spills	**All Spills**	**Total**
Oil Spills:				
Open-ended format		406	411	817
Dichotomous-choice format		393	397	790

Note: The goals for completed questionnaires were 400 for each open-ended alternative and 390 for each dichotomous-choice alternative (i.e., 65 for each of 6 bids).

Both Southlake and Lakeshore Malls are large, suburban, enclosed malls. Southlake Mall is located in the southern suburbs of Atlanta. It has 120 stores with four major anchor stores. Lakeshore is a smaller mall with two major anchors. It is located in Gainesville, Georgia, a more rural area north of Atlanta.

The recruiters used slightly different administration techniques at the two malls. In Southlake Mall, recruiters were stationed at one location on the upper level of the mall. The recruiters intercepted people and asked them if they would participate. If the respondent agreed, the respondent was escorted down a side corridor to a room and given the questionnaire. In Lakeshore Mall, the recruiters were allowed by the mall management to circulate throughout the mall to intercept respondents. Respondents then sat at benches in the mall and filled out the questionnaires using clipboards. On average, the surveys took 10 to 12 minutes to complete.

Table 3-4 summarizes the socioeconomic characteristics of the respondents for each mall. While most of the characteristics are quite similar across the malls, the Southlake Mall respondents are somewhat younger, have a higher income, and a much higher percentage of Southlake Mall respondents are males. Subsequent multivariate analysis indicated that the mall at which the respondent took the survey generally had no significant effect on the WTP responses, after controlling

Table 3-4. Characteristics of Respondents in Southlake and Lakeshore Malls

	Southlake Mall		Lakeshore Mall	
	Mean	Median	Mean	Median
Age (category in years)	30-39	20-29	30-39	30-39
Education (category)	some college	some college	some college	some college
Sex (% male)	65	NA	47	NA
Race (% non-white)	26	NA	22	NA
Income (category in $1,000)	35-50	35-50	25-35	25-35
People in Household (number)	2.9	3	3.2	3
Memberships in Environmental Organizations (number)	0.5	0	0.5	0

for socioeconomic differences. These results are documented in Chapter 5, when we report the full empirical analysis.

We chose the mall-intercept approach as a cost-effective way to test our hypotheses on people who could potentially be chosen for any CV survey. Our main goal is to have comparable samples for testing because the experiments are designed to compare differences across experimental treatments. They are not designed to develop damage estimates that are generalized to a specific population. Viscusi and O'Connor (1984) and Viscusi, Magat, and Huber (1985) have used mall surveys to conduct experimental tests in their risk studies. Additionally, mall respondents represent a broader group of the population compared with student populations, which have been used in several methodological studies (see Kealy, Dovidio, and Rockel, 1988; and Bergstrom, Stoll, and Randall, 1989).[3]

As discussed, each respondent received a randomly assigned survey version. This random assignment, coupled with the large sample sizes should result in subsamples with very comparable characteristics. Table 3-5 shows the distribution of sociodemographic characteristics by survey version. As expected, the distributions are very similar.

3.5 Cleaning WTP Responses

After survey data have been collected, some amount of data cleaning is necessary before any analysis begins. This process is especially important in CV surveys where there always are some invalid WTP responses. While it is easy to identify non-numeric responses, protest and outlier WTP responses are more difficult to identify. Nevertheless, identifying and excluding these invalid responses is important because leaving them in the analysis can have a large impact on the results. In the following sections, we first outline the various methods that have been used in past CV studies to identify invalid bids, and then we discuss the methods that we have employed.

[3] Mall intercept surveys have been deemed sufficiently reliable to be admitted as legal evidence (see McCarthy, 1991). The survey research literature states that an experiment needs only to be internally valid as long as the results will not be generalized to a larger population. Internal validity can be achieved sampling from any population when the sampling is random and the experimental and control groups are comparable. (See Sellitz, et al., 1970; and Babbie, 1979.)

Table 3-5. Distribution of Sociodemographic Characteristics for Each Design Alternative (%)

	Migratory Waterfowl			Open-Ended Oil Spills		Dichotomous-Choice Oil Spills	
	2,000 Birds	20,000 Birds	200,000 Birds	Small Spills	All Spills	Small Spills	All Spills
Age							
20–29	46	50	44	48	47	49	45
30–39	23	22	25	28	24	25	28
40–49	19	16	18	11	17	17	16
50–59	6	7	6	8	6	5	6
60–69	4	2	5	3	4	2	3
over 60	3	2	1	1	2	1	2
Education							
Some high school	6	7	7	6	6	4	3
High school diploma	29	25	33	27	29	27	26
Some college	31	36	28	33	35	35	34
College diploma	21	21	20	21	14	20	23
Some graduate school	5	4	4	5	5	4	5
Graduate degree	8	7	7	7	11	10	9
Sex							
Female	42	36	44	39	40	40	42
Male	58	64	56	61	60	60	58
Race							
Nonwhite	27	29	26	27	25	25	25
White	73	71	74	73	75	75	75
Income							
<15,000	11	12	12	12	10	10	10
15–25,000	19	16	20	19	20	24	15
25–35,000	17	21	21	20	19	17	21
35–50,000	24	23	22	22	23	21	26
50–65,000	14	13	9	10	14	12	15
65–80,000	7	7	8	8	7	7	8
80–100,000	4	4	2	4	4	4	2
>100,000	2	4	5	5	4	4	2

Methods Used in Past CV Studies

The last column of the table of nonuse studies in Appendix A indicates how several recent nonuse studies have addressed invalid responses. A review of these methods is hampered by the fact that many studies do not explicitly discuss how invalid bids are identified and treated. In this section we discuss alternative approaches used with open-ended question formats, which include iterative bidding games and payment cards as well as direct questions. These question formats usually produce a number of invalid WTP responses for several reasons. Some respondents reject the commodity being valued, the payment vehicle, or the facts or other aspects of the CV scenario. They may register their objections either by stating a $0 WTP amount, stating no WTP amount, or stating a value much larger than their actual WTP. Invalid responses also may be a consequence of respondents' not taking the survey seriously, not considering their income and expenses, or simple confusion about what is being asked. Additionally, some respondents find it too difficult to determine a dollar amount for the contingent commodity. Finally, some WTP amounts are considered outliers because they do not satisfy criteria defined by the researcher.

Several methods have been used to handle invalid WTP responses in past CV studies where the valuation question solicits a dollar response (e.g., open-ended, payment card, and bidding games). These procedures include:

- include all responses,
- reject protest zero responses based on answers to probe questions,
- reject positive bids that are greater than some specified percentage of income,
- trim some specified percentage of bids off both ends of the distribution, and
- reject outliers identified using statistical criteria.

Simply retaining all usable responses was used only in early CV studies (Bennett, 1984; Greenley, Walsh, and Young, 1981). In studies since 1984, researchers have realized that invalid responses can have a large impact on estimated mean and median WTP values and have taken steps to identify and eliminate them.

Using follow-up probe questions to identify protest zero responses is relatively common (Mitchell and Carson, 1984; Rowe et al., 1991; Smith and Desvousges, 1986b). Respondents who give a $0 response to the WTP question are asked for their primary reason for that response. If the answer indicates that the respondents indeed have some positive value but they reject the valuation mechanism, the responses are considered a protest.

Some respondents state relatively large WTP values. While some of these large amounts may accurately represent actual WTP, other large amounts clearly overstate actual WTP. One frequently used test of reasonableness is to compare the WTP amount with the respondent's income. If the amount is greater than some percentage of income, it is rejected as unreasonable. For example, Rowe et al. (1991) reject amounts in excess of 1 percent of the respondent's income. We have used a 25-percent screen in this study. Since the screen percentage is essentially arbitrary, using a 25-percent screen retains more large values than in some other studies.

Some prominent CV researchers appeal to the theory of robust statistics in using trimmed distributions (Carson, 1991). The researcher selects some percentage (usually 5 percent or 10 percent) and trims that percentage of the distribution from both the upper and lower end. This method removes the influence of the most extreme bids. One drawback of this and all other approaches is that model estimates may be affected by which observations are trimmed from the distribution, as well as the problems associated with trimming asymmetric distributions. (See McFadden and Leonard [1992] for more discussion.)

Deviant or inconsistent WTP amounts need not occur only in the tails of the distribution. The Belsley, Kuh, and Welsch (1980) method employs a statistical criterion for identifying observations that exert an undue influence on a regression equation for WTP. Those amounts then are removed as outliers. This method has been used in the past by Smith and Desvousges (1986b) and Reiling et al. (1989). We also use this technique to remove outlier observations from our open-ended data, as explained below.

Excluded Responses in Open-Ended CV Surveys

We use several of the above techniques to exclude responses in our open-ended data. The survey instrument includes follow-up probes, which ask respondents to explain the reason(s) for their answer. As noted above, respondents answer a different set of probes according to their WTP response. There is one set of probes for respondents giving zero or non-numeric bids, and one set for respondents giving positive bids.[4] We developed these probes from comments provided by pretest participants.

Question 6 in the migratory-waterfowl questionnaires (Question 8 in the oil-spills questionnaires) asks respondents who gave a $0 WTP response to indicate the most important reason for their response. In the migratory-waterfowl questionnaires we consider the following reasons to be indicative of protest responses:

- My household should not have to pay to protect these migratory waterfowl.
- There wasn't enough information for me to answer the question.
- Higher prices are not a good way to pay for protecting these migratory waterfowl.
- Wire-net covers would not be effective in protecting these migratory waterfowl.
- I could not determine a dollar amount for protecting these migratory waterfowl.

Respondents also were given the opportunity to write in a response if they felt that none of the stated responses represented their answer. We have classified some of these reasons as indicating protest $0 responses, such as "the oil companies should pay for these covers" and "the government should reallocate its budget to solve this problem." Comparable responses in the oil-spills survey also are classified as protests. Overall, these responses indicate that the respondent rejected the valuation exercise in some way.

Table 3-6 presents summary information on stated WTP from the migratory-waterfowl and oil-spills open-ended questionnaires. These questionnaires yield usable WTP data for about 70 percent of the sample. About 12 percent of the migratory-waterfowl respondents have a WTP of $0, while 59 percent have a positive WTP. Another 8 percent of these respondents gave a $0 WTP, but the follow-up probes reveal that these respondents are protestors. About 3 percent of the responses are identified as outliers. About one-sixth of the migratory-waterfowl respondents did not provide a numeric WTP response. Finally, 1 percent of the respondents gave WTP amounts that are greater than 25 percent of their income, or are $10,000 or more. We eliminate these responses from further analysis.

[4] To date we have not used the probe responses to exclude any positive WTP responses. We plan to explore this possibility in subsequent research.

Table 3-6. Summary Information on Open-Ended WTP Responses

	Migratory Waterfowl		Oil Spills (Open-Ended)	
	#	%	#	%
Included				
$0	146	12	73	9
>$0	709	59	484	59
Excluded				
$0 (Protest)	95	8	102	12
>$0 (Income Screen)	18[a]	1	8[b]	1
>$0 (Outliers)[c]	31	3	19	2
Non-numeric responses	206	17	131	16
Total	1,205	100	817	100

[a] Includes amounts greater than 25 percent of the respondent's income and a $12,000 amount.
[b] Includes amounts greater than 25 percent of the respondent's income and a $10,000 amount.
[c] Outliers identified using the Belsley, Kuh, and Welsch (1980) technique.

As shown in Table 3-6, the proportions of included and excluded WTP responses in the open-ended oil-spills survey are very similar to the migratory-waterfowl survey. There are more excluded $0 responses than included $0 responses in the oil-spills survey, but the proportions in the other categories are remarkably similar to the migratory-waterfowl WTP responses.

In total, we identified 31 outliers in the migratory-waterfowl WTP responses and 19 outliers in the oil-spills WTP responses. Appendix Tables E-1 and E-2 provide profiles of these outliers. As shown in these tables, all but two of the outlier observations have a stated WTP of $1,000 or more.[5]

Our extensive pretesting of the survey instruments appears to have yielded effective questionnaires.[6] Only 6 respondents did not complete their questionnaire. However, we have a sizeable proportion of excluded responses (29 percent for birds, and 32 percent for oil spills), which may cause some concern. Some recent

[5] We also completed a probit analysis on the probability of a respondent being classified as an outlier. The analysis showed that respondents in their 20s, having higher incomes, and belonging to more environmental organizations are more likely to be outliers.

[6] Only 6 respondents failed to answer any of the sociodemographic questions.

CV studies have had lower percentages of excluded WTP responses. For example, Carson's (1991) study of visibility reduction in the Grand Canyon excluded about 10 percent of the responses. However, the Grand Canyon study uses in-person interviews and involves a well-known, unique natural resource with some use value. In contrast, our instruments are self-administered and primarily involve nonuse values for less prominent natural resources. The Rowe et al. (1991) study of the *Nestucca* oil spill off the coast of Washington State excluded about 40 percent of the observations. That study used a mail survey to measure WTP values for a non-unique commodity, making it more comparable to our commodities. Also, the Desvousges, Smith, and McGivney (1983) study of the Monongahela River resulted in about 19 percent protest zeros. Again, that resource is a non-unique commodity.

In the remainder of this report, we provide results based on the included open-ended data, which we call the censored data. These data eliminate

- protest $0 responses,
- responses that fail the 25 percent-of-income criterion,
- WTP responses greater than or equal to $10,000, and
- outliers based on the Belsley, Kuh, and Welsch technique.

In other analyses presented in the Appendices G through H, we use several other versions of the data:

- screened data: does not eliminate outliers identified with the Belsley, Kuh, and Welsch technique; and
- trimmed data: 5 percent and 10 percent trims performed on the screened data.

Excluded Responses in Dichotomous-Choice CV Surveys

Unlike the open-ended CV format in which respondents must determine their WTP for the contingent commodity, the dichotomous-choice CV format only requires respondents either to agree or disagree to pay a specified amount for the contingent commodity. Furthermore, most dichotomous-choice CV surveys do not explore respondents' reasons for their Yes or No responses, which means that protestors are not identified. Consequently, the only excluded responses come from respondents who do not answer the CV question and respondents who give both a

Yes and No as their response to the CV question. As shown in Table 3-7, we excluded only 2 percent (16 of 790) of the responses to our dichotomous-choice oil-spills survey. In contrast, 32 percent (260 of 817) of the responses to the open-ended oil-spills survey are excluded (see Table 3-6). Although our dichotomous-choice surveys contain probe questions, we have not used these probe responses to eliminate any WTP responses. We plan to incorporate the probe responses in future research.

Table 3-7. Summary Information on WTP Responses to Dichotomous-Choice Oil-Spills Questionnaires

	Number	Percent
Included		
Yes	394	50
No	380	48
Excluded		
Other	16	2
Total	790	100

3.6 Summary

This chapter summarizes the design, development, and administration of our surveys. During the development process we employed well-established techniques to ensure effective questionnaires. We have tried to minimize any problems with our survey instruments that might confound the tests of our hypotheses about accuracy.

After survey data are collected, some amount of data cleaning is necessary. Although this process is common in CV surveys, no unambiguous standard exists for data cleaning. Generally, we use conservative criteria that we think are consistent with good CV practice. However, we also report the results of using two alternative methods for excluding WTP responses in appendices. As explained later, these alternative methods do not substantively change any of the results presented in Chapters 5, 6, and 7.

CHAPTER 4

A Conceptual Framework for Estimating Nonuse Values

4.1 Background

The rationale for using CV for damage assessments is that environmental accidents often affect the quality of environmental services for which there are no observable market values. CV provides a means of obtaining monetary measures of welfare changes in order to evaluate and justify appropriate compensation payments. To interpret CV values as welfare measures, the analyst must make several important assumptions:

- Respondents have well-defined, well-behaved preferences for the damaged environmental resource services.
- Respondents are rational in the sense of consistently choosing among hypothetical alternatives to maximize utility.
- Respondents' preferences are exogenous to the process that caused the damage and to the CV instrument itself.
- Respondents' expressed values for the damaged resources as described by the CV instrument are the same as what their revealed values would be in an actual choice situation.

NRDAs primarily are concerned with the total value of damages associated with a specific environmental accident. In contrast, this study focuses on whether different levels of damage produce expressed values that are consistent with economic theory. If repeated tests of this hypothesis under controlled experimental conditions consistently reject theoretical validity, we must conclude that one or more of the underlying assumptions does not hold.

In principle, total contingent values can be obtained by drawing an appropriate random sample of the population and extrapolating the observed distribution of values to that population. Concern about the influence of a few large values on mean WTP have led several authors to employ trimmed means or medians as measures of central tendency (see Carson, 1991). McFadden and Leonard (1992) criticize such attempts to minimize the influence of extreme bids. They argue that trimmed means are biased estimates of the mean for asymmetric distributions. Medians are inconsistent with total damage values based on classical welfare theory.

Furthermore, directly extrapolating observed values does not relate values to personal characteristics of respondents, does not incorporate the preference and behavior assumptions listed above, and limits the analyst's ability to compare results across different CV experiments. Therefore CV studies often specify a parametric model that relates responses to open-ended or dichotomous-choice questions to such explanatory variables as age, income, and measures of environmental concern. Unfortunately empirical models reported in the literature are not always consistent with utility theory (Bishop, Heberlein, and Kealy, 1983; Bowker and Stoll, 1988; and Boyle and Bishop, 1988). The relationship of resulting estimates to welfare values is uncertain in such cases.

When the estimation procedure is consistent with utility theory, multivariate models make it possible to test a variety of hypotheses. For example, we would expect wealthier respondents and people with pro-environmental attitudes to be willing to pay more, other things held constant. However, multivariate models also pose the obvious problem of having to identify the particular structure that best fits the data. Parametric estimates of mean WTP also are sensitive to the shape of the tails of the assumed distribution. This chapter proposes a general utility-theoretic model of nonuse values and discusses its implications for estimating and interpreting multivariate WTP models for both open-ended and dichotomous-choice formats.

4.2 A Utility-Theoretic Model of Natural Resource Values

In order to test whether expressed nonuse values are consistent with economic theory, we derive a utility-theoretic model of nonuse values. The model employs a flexible, Box-Cox functional form for the utility function that incorporates popular empirical specifications of WTP as special cases.[1] Suppose an individual faces a choice between a quantity of an environmental commodity not being available or a quantity being available at a fixed charge t and marginal price p. An example could be a park whose maintenance is financed partly from general tax revenues and partly from user fees. Assume a flexible functional form for indirect utility when the quantity is not available:

$$V(y;0) = \frac{y^{1-\alpha} - 1}{1 - \alpha} \tag{4.1}$$

where y is income. When the specified quantity of the commodity is available, indirect utility is an additively separable stochastic function

$$V(y - t,p;E) = \frac{(y - t)^{1-\alpha} - 1}{1 - \alpha} + m(\Theta,x)e^{-\tau p} \tag{4.2}$$

where E is the available amount of the environmental commodity, τ is a parameter, $m = m(\Theta,x)$ is a random variable with probability density function $f(m;x,\Theta)$, x is a vector of socioeconomic characteristics and preference indicators, and Θ is a vector of parameters.

If the commodity provides use value, then Roy's identity yields the demand function

$$E = m\tau(y - t)^{\alpha} e^{-\tau p} \tag{4.3}$$

Thus α is the income elasticity of demand and τ is the own-price elasticity. However, if the commodity provides only nonuse value, p is undefined, $\tau = 0$, and E is constant

[1] This exposition is basically consistent with several other theoretical models, including Hanemann (1984), Hoehn and Randall (1987), and McConnell (1990) among others. Except for notational differences, it follows McFadden and Leonard (1992).

for all consumers. In the absence of such variation the information necessary to recover the demand parameters usually is obtained from a CV survey.[2]

4.3 Estimating WTP Values From Open-Ended Data

An open-ended CV question directly asks respondents their maximum WTP for the commodity described in the survey instrument. Utility maximization presumes that the open-ended question will elicit a value t=W that equates indirect utility with and without the commodity.[3] Therefore willingness to pay W satisfies[4]

$$m = G(W,y;\alpha) = \begin{cases} W & \text{if } \alpha = 0 \quad (4.4a) \\ [y^{(1-\alpha)} - (y - W)^{(1-\alpha)}]/(1-\alpha) & \text{if } 0 < \alpha < 1 \quad (4.4b) \\ -\,Ln(1 - W/y) & \text{if } \alpha = 1 \quad (4.4c) \end{cases}$$

It follows that W has a probability density function

$$f(W;x,\Theta) = f(m;x,\Theta)\ dm/dw = \frac{f\,[G(W,y;\alpha);x,\Theta]}{(y-W)^{\alpha}} \tag{4.5}$$

For general case (4.4b) solving for W gives

$$W = y - [y^{(1-\alpha)} - (1-\alpha)m)]^{1/(1-\alpha)} \tag{4.6}$$

Except for the special case of $\alpha = 0$, W is a nonlinear function of the stochastic variable m for each respondent i, so the mean of $E(W_i)$ will not in general equal the sample mean of W_i.[5] This implies that the sample mean of open-ended CV responses could differ from the expected value of WTP derived from a utility-theoretic model.

---

[2] Larson (1991) has proposed examining so-called indirect uses such as reading magazines and watching television documentaries as a behavioral basis for estimating nonuse values.

[3] W is the true (but stochastic) Hicksian compensating surplus. The Hicksian equivalent surplus (willingness to sell (WTS) in contrast to WTP) can be derived by using V(y+t,p;E) in Equation 4.2. It is beyond the scope of this monograph to discuss possible discrepancies between empirical measures of value obtainable from CV surveys and the conceptually correct measure for compensable damages. (For example, see Cummings, Brookshire, and Schulze, 1986; Hoehn and Randall, 1987; Hanemann, 1991; and Carson, Flores, and Hanemann, 1992.)

[4] We are implicitly assuming that W is less than some discretionary portion of total income that is available for all environmental purposes. If W exceeds that fraction of income, $m = \infty$.

[5] For example, when $\alpha = 1$, W = y-yexp(m). The sample mean of W_i is a function of $\exp(\Sigma m_i/n)$, which is not equal to the mean expected value of W_i, which is a function of $\Sigma \exp(m_i)/n$.

W is always non-negative if the commodity is a good for all respondents. This constraint implies that W is censored at zero, and $m \geq 0$. If m is normally distributed with mean $x\beta$ and variance σ^2, $m = x\beta + \sigma u$, where u is a standard normal disturbance, then tobit is an appropriate estimator for the two polar cases (4.4a) and (4.4c).[6] If α is not constrained to be zero or unity, or if normality is not assumed, then other maximum likelihood techniques must be applied.

In Chapters 5 and 6 we report tobit estimates for the linear and natural log cases, which are commonly encountered in the CV literature. These estimates then are used to evaluate the theoretical validity of parameter estimates and to derive variance estimates to evaluate reliability. Appendix F reports some results for more general specifications that allow for flexible α values and non-normality.

4.4 Estimating WTP Values From Dichotomous-Choice Data

The dichotomous-choice or referendum question format confronts respondents with a randomly assigned bid and asks them whether they would be willing to pay the stated amount. Since respondents do not reveal their maximum WTP in this kind of survey, it is necessary to estimate WTP from the observed pattern of responses. The econometric theory used to analyze this kind of problem was developed in the 1970s for such applications as travel-mode choice (Domencich and McFadden, 1975), labor markets (Heckman and Willis, 1976), appliance purchases (Hausman, 1979), and new products (Beggs, Cardell, and Hausman, 1981). Bishop and Herberlein's (1979) study of goose-hunting permits was the first use of dichotomous-choice data to value a natural resource.

When confronted with a bid t, the respondent will accept the bid if $t \leq W$ with a probability

$$P(\text{Yes} \mid \Theta;t,x,y) = 1 - F[G(t,y;\alpha),x,\Theta] \tag{4.7}$$

[6] Maddala (1983) discusses several approaches to predicting values in the tobit model. The predicted values ($x\beta$ in the linear model) represent the predicted values of the unobserved "true" WTP. When the value of this latent variable is less than or equal to zero, the tobit model assumes that we observe zero stated WTP. Assigning values of zero to such observations yields an estimate of the mean and median WTP. Alternatively, we can predict the stated WTP reflecting the error inherent in the observed values. We report means and medians based on the first of these two measures in Chapters 6 and 7.

where F[•] is the cumulative distribution function. If F[•] equals the cumulative normal distribution function, Φ[•], then

$$F[G(t,y;\alpha),x,\beta,\sigma] = \begin{cases} \Phi[t - x\beta)/\sigma] & \text{if } \alpha = 0 \quad (4.8a) \\ \Phi\left[\dfrac{[\{y - (y - t)^{(1-\alpha)}\}/(1 - \alpha)] - x\beta}{\sigma}\right] & \text{if } \alpha < 1 \quad (4.8b) \\ \Phi[\{-\text{Ln}(1 - t/y) - x\beta\}/\sigma] & \text{if } \alpha = 1 \quad (4.8c) \end{cases}$$

Several studies have shown that the above specification is merely a reparameterization of logit and probit models. (See Hanemann, 1984; Cameron, 1988; Cameron and James, 1987; McConnell, 1990; and Patterson and Duffield, 1991.) For the probit case,

$$P(\text{Yes}) = P[V(y;0) \leq V(y - t;E)] = 1 - \Phi(-x\gamma - bt) \qquad (4.9)$$

so that $\sigma = -1/b$ and $\beta = -\gamma/b$. Thus σ and β can be estimated directly by maximum likelihood or can be derived from probit estimates.[7]

Figure 4-1 illustrates the m = G(•) function when α = 0 and m is normally distributed, while Figure 4-2 is the corresponding P(Yes) function. The frequency distribution of WTP in Figure 4-3 is simply the derivative of P(Yes) or the change in P(Yes) at various bid levels.

The mean WTP is the area under the P(Yes) function, while the median WTP is the value at which m = 0 or the estimated probability of answering Yes equals 0.5.[8] Note that it is possible for the estimated median to be negative. If the probability

[7] For the linear specification, Equation 4.8 has the advantage that the estimated coefficients to be interpreted as the derivatives of WTP with respect to each independent variable. However, Cameron's (1988) applications use 1n(bid) to constrain WTP to be strictly positive. This functional form precludes zero values and creates potential problems of choosing an upper integration bound. Appendix F employs a mixed distribution to accommodate zeros and demonstrates the sensitivity of calculated means to choice of integration bound.

[8] Johansson, Kriström, and Miler (1989) point out that if the selected functional form does not preclude negative WTP values, the mean of the estimated distribution is the area under P(Yes) from zero to infinity *minus* the area under 1-P(Yes) from minus infinity to zero. These areas are shaded in Figure 4-2.

Figure 4-1. Preference Function: $m = x\beta + \sigma\mu$

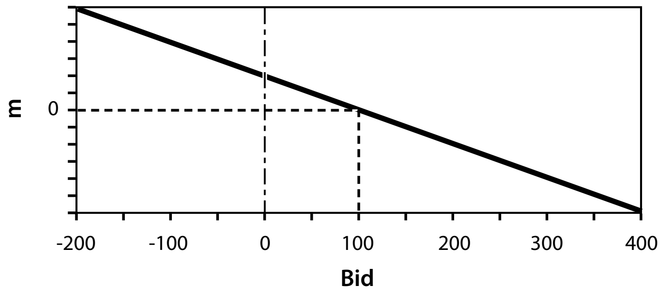


Figure 4-2. Probability of Yes Response

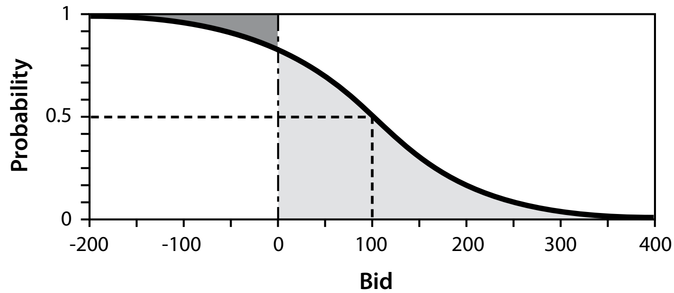
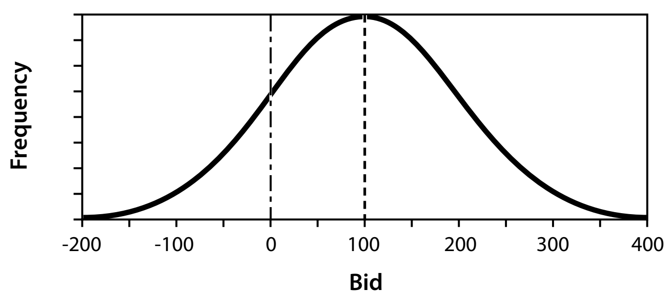


Figure 4-3. Distribution of WTP



of accepting a very small bid is sufficiently small, the intercept of the probability gradient could be less than 0.5 and the median would lie to the left of zero. Such a result is reported by Bowker and Stoll (1988).

4.5 Model Specification

Equations 4.4 and 4.8 imply that contingent values should be estimable as a stochastic function of a vector of typical demand-shift variables x. The rationale for choosing the variables and functional form for a particular model either can be guidance from economic theory or statistical criteria. Economic theory often does not provide specific guidance on variable selection, apart from income and prices. However, most previous CV studies suggest that WTP for environmental commodities should depend on several explanatory variables. (For example, see Mitchell and Carson, 1989; and Freeman, 1990.) Among these are respondents'

- knowledge of the contingent commodity,
- environmental attitudes and opinions, and
- socioeconomic characteristics (such as age, education, race, and sex).

Additionally, the offered bid is included as an explanatory variable for probit or logit specifications.

The selection of a functional form for CV dichotomous-choice models is controversial (see Boyle, 1990, for a summary of this literature). Hanemann (1984) proposes two utility-theoretic functional forms that correspond to Equations 4.8a and 4.8c. Other researchers have used specifications that are not derived from indirect utility functions. However, these ad hoc alternatives often perform better statistically than those based on utility theory (Boyle and Bishop, 1988 and Bowker and Stoll, 1988).

Some CV studies have found that income is a significant determinant of WTP. Nevertheless, Hanemann (1984) and McFadden and Leonard (1992) argue that utility-theoretic specifications of m should not include income in the linear case, and only as ln(1-t/y) in the log case because conventional theory requires that consumers treat a fixed charge simply as an adjustment to income. However, excluding income from x assumes that the marginal utility of income is constant between the two utility states under consideration. The literature on option value (see, for example, Cook and Graham, 1977, and Smith, 1984) shows that marginal utility of income is constant only for perfectly replaceable commodities. Making m a function of income allows testing for perceived uniqueness.

Estimates reported in Chapters 5, 6, and 7 are based on the theoretically consistent linear and log specifications derived above. Income is included as an explanatory variable in order to check for constancy of marginal utility of income. Estimates for some general functional forms that allow for flexible α are summarized in Appendix F.

4.6 Summary

We have employed utility theory to motivate estimating multivariate models that identify and quantify economic relationships in CV data. The theory imposes constraints on admissible functional forms and explanatory variables. It also provides guidance on the mechanics of estimating parameters from open-ended and dichotomous-choice question formats. The remainder of the monograph reports and interprets our empirical results using the approach presented above.

CHAPTER 5

Tests of Theoretical Validity

5.1 Background

Theoretical validity is established when an estimate conforms to a predetermined theoretical construct. For the current application, we investigate whether estimated contingent values conform to a fundamental premise of the theory of consumer demand. That is, for any item yielding positive marginal utility, additional amounts of the commodity will increase an individual's total utility. This result follows directly from the first assumption cited at the beginning of Chapter 4: respondents have well-defined, well-behaved preferences for injured environmental resources. Thus, theoretically valid CV estimates of nonuse values should increase as the level of injury increases, and such increases should be detectable for a range of injury that actually occurs.

We conduct three independent tests of this hypothesis. First, with the migratory-waterfowl data we expect the estimated mean value for preventing 200,000 bird deaths to exceed the comparable estimate for preventing 20,000 bird deaths, which would exceed the mean for preventing 2,000 bird deaths. Secondly, we expect the open-ended oil-spills survey to reveal that the value of reducing environmental effects from all oil spills exceeds the value of such protection from only spills less than 50,000 gallons. The dichotomous-choice estimates for the oil-spills scenarios provide the third test of this hypothesis.

A review of studies estimating values for which nonuse is a major component reveals little effort to establish the validity of marginal value estimates. These studies have focused on establishing that people hold values that go beyond use. Greenley, Walsh, and Young (1981), for example, try to estimate values for water quality in the South Platte River for existence, bequest, and option values that go beyond current use. Likewise, Brookshire, Eubanks, and Randall (1983) simply try to estimate existence values for grizzly bears and big horn sheep in Wyoming, in addition to estimating values for hunting and viewing.

The more recent trend has been to estimate option prices which include all components of value, including use and nonuse (Boyle and Bishop, 1987; Carson, 1991; and Rowe et al., 1991). However, unless nonuse values are a major component of total value, this approach obscures any potential irregularity in the nonuse components. In contrast, Mitchell and Carson (1984) and Smith and Desvousges (1986b) both estimate option prices for boatable, fishable, and swimmable water quality, and show that nonuse values comprise a significant component of the valuation estimates. These applications are akin to our oil-spill experiment and, in theory, offer an opportunity to test a theoretical validity hypothesis similar to the one we propose in this chapter. That is, do respondents hold higher values for sequential improvements in water quality? The ability of these studies to properly test the theoretical validity of their estimates is compromised, however, because of their multiple objectives.

To examine theoretical validity in this chapter, we conduct four types of statistical tests for each of the two open-ended surveys. The first two comparisons use nonparametric tests of responses to the open-ended data: Wilcoxon rank sum and permutations tests. These tests are used because of the apparent nonnormality of the WTP data. Nonparametric tests make no assumptions about the underlying distributions of the data being tested. The third and fourth tests use estimated equations to document the relationships between WTP and selected explanatory variables, as discussed in Chapter 4. Two types of statistical tests are performed on these estimates. The first is a test of differences between estimated mean values using z-statistics. The second is a test of equality between estimated vectors of coefficients in the equations using likelihood-ratio tests. For the dichotomous-choice oil-spill data, we also perform z-tests on means estimated with equations, as well as likelihood-ratio tests. In addition, we compare the distributions of Yes/No responses using a chi-square test. This final set of tests allows us to investigate whether our results change with the use of a different question format.

5.2 Migratory Waterfowl

Table 5-1 shows the univariate statistics of the WTP responses for each of the three versions of the migratory-waterfowl questionnaire. All of these descriptive statistics are fairly consistent across the versions. For example, using the censored data the means range from $78 to $88, which is a relatively small spread given the standard deviations of $132 to $187. The median is $25 for all three versions, and the range of values is quite similar for all three questionnaires. Nevertheless, the univariate statistics must be interpreted with caution. As shown by the Shapiro-Wilk statistic, these data are not normally distributed. Given the complex nature of WTP responses, therefore, more sophisticated analyses are necessary to test for significant differences across versions.

Figure 5-1 presents the frequency distributions of WTP responses for each of the three levels of migratory-waterfowl protection: 2,000, 20,000, and 200,000 birds. (To conserve space, this figure uses a base-10 logarithmic scale to present the WTP responses.) The similarities among the distributions are very striking. The response patterns have similar peaks and generally follow the same distribution.

Table 5-2 presents the results of the nonparametric tests of differences in the WTP distributions of the three levels of migratory-waterfowl deaths. We performed two nonparametric tests: the Wilcoxon rank sum (WRS) test and the permutations test.

Table 5-1. Univariate Results for Three Migratory-Waterfowl Questionnaires: Censored Data Sets

	2,000 Birds (WTP $)	20,000 Birds (WTP $)	200,000 Birds (WTP $)
Mean	80	78	88
Standard Deviation	187	132	166
Median	25	25	25
Mode	0	100	100
Range	0–1,550	0–1,000	0–1,000
Shapiro-Wilk statistic[a]	0.43	0.60	0.54
N	288	286	281

[a] This test statistic indicates that these distributions are not normal.

Table 5-2. Nonparametric Tests of Differences in WTP Distributions: 2,000 Birds, 20,000 Birds, and 200,000 Birds[a]

Version Comparison	Monte Carlo p-value Estimates
2,000 Birds vs. 20,000 Birds	
Wilcoxon Rank Sum Test	0.178
Permutations Test	0.425
20,000 Birds vs. 200,000 Birds	
Wilcoxon Rank Sum Test	0.457
Permutations Test	0.185
2,000 Birds vs. 200,000 Birds	
Wilcoxon Rank Sum Test	0.145
Permutations Test	0.288

[a] Test statistics are Monte Carlo p-value estimates at 99 percent level of confidence.

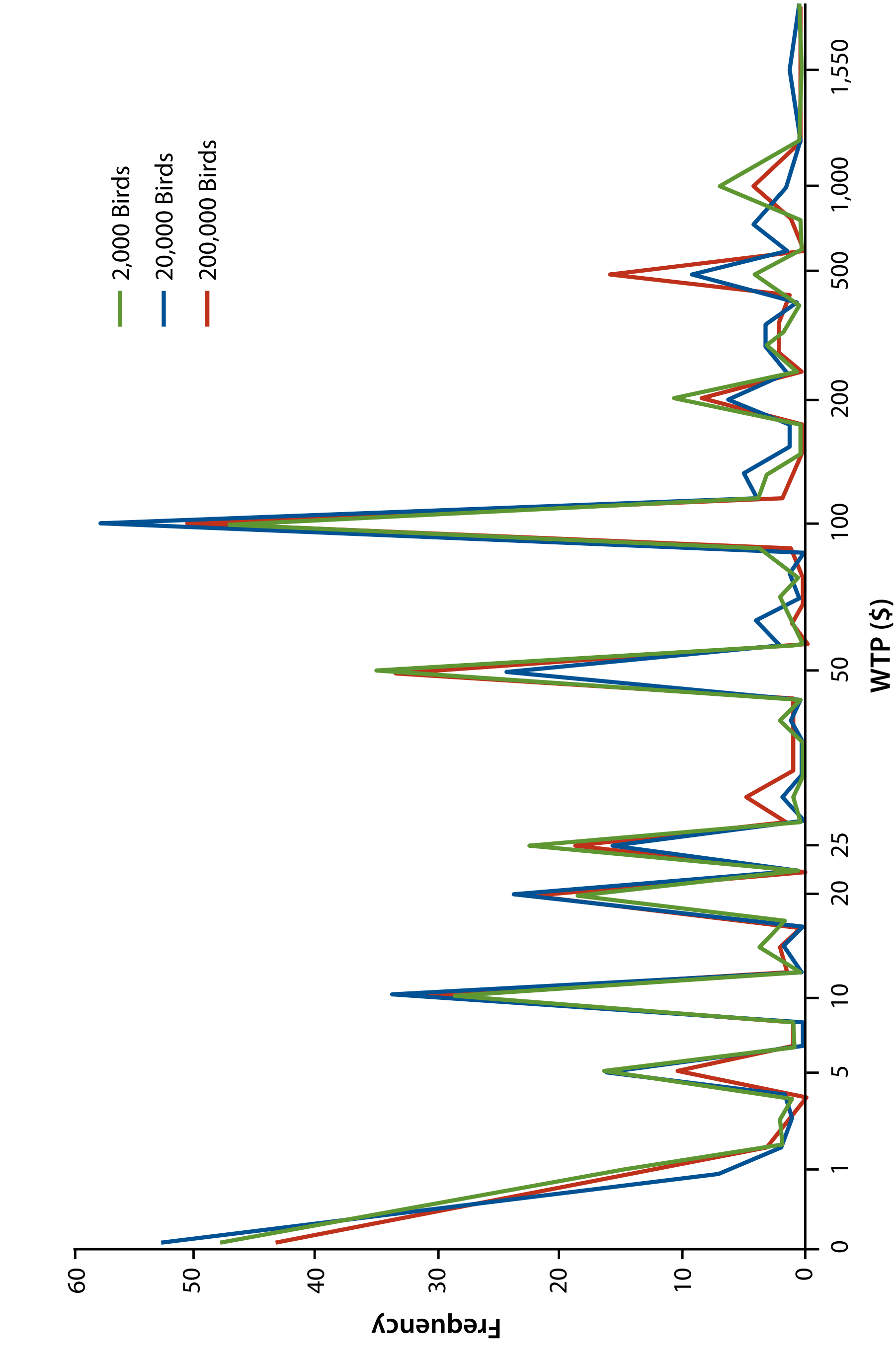


Figure 5-1. WTP Frequency Distribution: Migratory Waterfowl

The WRS test is a standard nonparametric technique to determine if two populations have the same distribution.[1] The output from this test is a probability value (p-value). For our data, this p-value is estimated using a Monte Carlo technique. The p-values from this test are similar to p-values from standard parametric tests. They represent the statistical significance of the test, so a p-value of 0.10 indicates that the test of no difference in the distributions is statistically significant at the 10-percent level.

The permutations test is an alternative nonparametric technique for testing for differences in distributions.[2] As with the WRS test, the output from the permutations test is an estimated p-value generated from a Monte Carlo technique. This test is very sensitive to outliers, so the WRS test probably is more useful for our data.

The numbers in Table 5-2 are the estimate of the p-value. The null hypothesis of no differences in the distributions cannot be rejected for any of the three comparisons using either the WRS test or the permutations test. These tests show that the distributions of WTP values are not statistically different. This result is inconsistent with the theoretical validity of the estimated values because protecting a larger number of migratory waterfowl in the Central Flyway does not produce significant changes in the distributions of responses.

To test further for theoretical validity, we use a z-test to evaluate the significance of differences in mean WTP for the three migratory-waterfowl questionnaires. As discussed in Chapter 4, we modeled the open-ended data using tobit models.

---

[1] The test is performed by combining the observations from the two samples and sorting them in ascending order. The value of each observation is then replaced with its rank (that is, the smallest observation is given the rank of 1, the next is given 2, and so on). By using the ranks instead of the actual values, the test is not affected by outliers or other distribution characteristics. The test statistic is then derived from the sum of the ranks of the observations from one of the two samples. If that test statistic is sufficiently different from one-half the sum of all the ranks, then the test rejects the null hypothesis that the populations are the same.

[2] In this test, instead of using the ranks to calculate the test statistic, the actual values, or "general scores" are used. This test compares the average score (or mean) from each of the two samples, making this test similar to the standard z-test. However, the permutations test does not require any assumptions about the distribution of the underlying data, while the z-test requires normally distributed data. Instead of comparing the data to a standard normal distribution, an underlying distribution is generated that is specific to the given set of data.

The specification of the models used for the hypothesis tests pose a problem. In order to avoid variation in estimates across treatments due to differences in model specification, we use the same specification for all the treatments. In a few cases, the standard specification does not represent the "best" possible fit or the set of explanatory variables with the most statistically significant coefficients. However, differences between the standard and best specifications have very little effect on estimated WTP. Using best specifications instead of the standard specification does not change our results. For comparison purposes, we also report specifications using no explanatory variables and others using a large number of explanatory variables, including income.

We estimate the two polar cases discussed in Chapter 4. The linear case assumes income elasticity is equal to zero, while the log case assumes income elasticity is equal to one. (See Appendix F for estimates of the general case.) Table 5-3 defines the explanatory variables included in the models. Tables 5-4 through 5-6 report the estimated tobit coefficients for the three versions of the migratory-waterfowl survey for three different model specifications. SIGMA is the estimate of σ defined in Equations 4.2 and 4.4. First, we discuss the full models presented in Tables 5-4 through 5-6. These models include variables from all parts of the survey.

Despite the inclusion of variables that economic theory and common sense indicate should be related to expressed WTP, the full models have very few statistically significant coefficients. Although coefficient estimates rarely are significant, coefficient signs generally are consistent with expectations. The unexpected negative signs on the variable for the number of threats for which respondents have high knowledge (K_COUNT) and an environmental attitude variable (O_EXIST) and the unexpected positive signs on whether respondents thought the covers would not affect the waterfowl population significantly (O _ AFFNA) are not statistically significant. The MALL coefficient is significant for the linear 200,000-birds case, suggesting that respondents at Southlake Mall had higher WTP, holding other variables constant. Since assignment of treatments was carefully randomized between and within malls, and none of the other treatments reveal mall differences, this result appears to be a statistical artifact. INCOME is insignificant for the linear case in all three treatments, but significant and unexpectedly negative for the log case. Neither of these refutes our assumption of constant marginal utility of income, as described in Chapter 4.

Table 5-3. Description of Explanatory Variables in Tobit Analysis of Migratory-Waterfowl Questionnaires

Variable	Description
MALL	Dummy = 1 for Southlake Mall (0 for Lakeshore Mall)
READR	Number of times the respondent had read or heard about issues involving migratory waterfowl in the 6 months prior to the survey (Q.1)
K-COUNT	Number of threats to migratory waterfowl in the Central Flyway for which the respondent has high knowledge (Q.3)
O_EXIST	Dummy = 1 for respondents who disagree or strongly disagree that people have a right to change the environment to meet their needs ***and*** agree or strongly agree that people should preserve the environment at all costs (Q.7)
O_DEADD	Dummy = 1 for respondents who agree or strongly agree that any death of migratory waterfowl is a serious problem (Q.8)
O_AFFNA	Dummy = 1 for respondents who agree or strongly agree that covering waste-oil holding ponds will not significantly affect the migratory-waterfowl population in the Central Flyway (Q.8)
P_HMWA	Dummy =1 for respondents who have hunted migratory waterfowl anywhere in the United States (Q.9)
P_BIRDA	Dummy = 1 for respondents who have bird-watched anywhere in the United States (Q.9)
AGEYR	Age of respondent (midpoint of categories in Q.10, with 75 for the highest category)
NORGS	Number of environmental or conservation organization memberships in the respondent's household (Q.16)
INCOME	Respondent's household income (Q.14)

Note: Specific questions are referenced following the description. For example, (Q.8) refers to question 8 of the survey.

Table 5-4. WTP Models Based on Tobit Analysis: 2,000-Birds Version (t-statistics in parentheses)

	Linear Model			Log Model		
	Constant Only Model	Full Model	Final Model	Constant Only Model	Full Model	Final Model
INTERCEPT	54.86*** (4.04)	-6.72 (0.12)	13.00 (0.29)	0.002* (1.91)	0.003 (1.28)	0.001 (0.44)
READR		3.74 (0.53)	6.77 (1.02)		0.000 (0.03)	0.000 (0.11)
O_DEADD		64.09** (2.14)	62.26** (2.20)		0.002 (1.11)	0.002 (1.38)
P_HMWA		94.96*** (2.68)	112.48*** (3.32)		0.004* (1.78)	0.004*** (3.26)
AGEYR		-0.92 (0.82)	-0.91 (0.82)		-0.000 (0.39)	-0.000 (0.55)
MALL		4.23 (0.15)			0.001 (0.34)	
K_COUNT		7.39 (0.53)			-0.000 (0.16)	
O_EXIST		-35.84 (1.19)			-0.001 (0.31)	
O_AFFNA		-23.13 (0.78)			-0.001 (0.66)	
P_BIRDA		34.62 (1.17)			0.001 (0.48)	
NORGS		16.81 (1.13)			0.001 (0.63)	
INCOME		3.49 (0.61)			-0.001* (1.93)	
SIGMA	216.39*** (20.90)	207.77*** (20.69)	209.56*** (20.69)	0.009*** (39.70)	0.008*** (30.28)	0.008*** (34.17)
N	272	267	267	272	267	267
Pseudo-R^2	0.02	0.07	0.08	0.02	0.05	0.06
Estimated mean	55	58	57	73	61	76
Estimated median	55	41	50	77	38	43

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 5-5. WTP Models Based on Tobit Analysis: 20,000-Birds Version (t-statistics in parentheses)

	Linear Model			Log Model		
	Constant Only Model	Full Model	Final Model	Constant Only Model	Full Model	Final Model
INTERCEPT	59.12*** (6.43)	53.56 (1.53)	74.35** (2.48)	0.002** (2.39)	0.005 (1.56)	0.004 (1.17)
READR		5.76 (1.37)	10.56*** (2.57)		0.000 (1.11)	0.001*** (3.19)
O_DEADD		20.89 (1.16)	38.87** (2.24)		0.001 (0.34)	0.002 (1.19)
P_HMWA		24.88 (1.14)	42.49** (2.06)		-0.000 (0.09)	0.001 (0.30)
AGEYR		-2.28*** (3.09)	-2.04*** (2.73)		-0.000 (-1.61)	-0.000 (1.56)
MALL		-2.90 (0.14)			-0.000 (0.27)	
K_COUNT		22.65*** (2.64)			0.001 (1.58)	
O_EXIST		23.22 (1.32)			0.003 (1.45)	
O_AFFNA		2.10 (0.11)			-0.001 (0.58)	
P_BIRDA		22.52 (1.23)			0.002 (0.92)	
NORGS		11.08 (1.20)			0.001 (0.61)	
INCOME		3.45 (1.02)			-0.001* (1.66)	
SIGMA	148.53*** (20.90)	127.66*** (20.71)	131.35*** (20.56)	0.007*** (87.81)	0.008*** (29.81)	0.008*** (41.66)
N	279	274	271	279	271	271
Pseudo-R^2	0.02	0.14	0.11	0.02	0.07	0.05
Estimated mean	59	59	58	69	74	74
Estimated median	59	52	56	68	35	42

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 5-6. WTP Models Based on Tobit Analysis: 200,000-Birds Version (t-statistics in parentheses)

	Linear Model			Log Model		
	Constant Only Model	Full Model	Final Model	Constant Only Model	Full Model	Final Model
INTERCEPT	71.03*** (6.11)	6.87 (0.15)	67.45* (1.66)	0.003 (1.43)	0.003 (0.44)	0.002 (0.27)
READR		13.13** (2.09)	13.04** (2.13)		0.001 (1.46)	0.001* (1.68)
O_DEADD		25.30 (0.99)	43.54* (1.77)		0.002 (0.65)	0.004 (1.26)
P_HMWA		63.00** (2.18)	66.97** (2.36)		0.005 (1.25)	0.005 (1.52)
AGEYR		-1.61 (1.60)	-1.70* (1.69)		-0.000 (0.47)	-0.000 (0.76)
MALL		61.35*** (2.59)			0.004 (1.20)	
K_COUNT		-3.49 (0.29)			-0.002 (0.79)	
O_EXIST		14.62 (0.57)			0.002 (0.49)	
O_AFFNA		16.56 (0.62)			0.001 (0.27)	
P_BIRDA		27.42 (1.11)			0.003 (0.92)	
NORGS		17.63 (1.31)			0.001 (0.91)	
INCOME		2.68 (0.59)			-0.001* (1.83)	
SIGMA	186.76*** (21.17)	181.20*** (20.89)	184.55*** (20.85)	0.016*** (56.93)	0.016*** (27.40)	0.016*** (33.97)
N	272	264	264	272	264	264
Pseudo-R^2	0.01	0.09	0.08	0.02	0.04	0.04
Estimated mean	71	73	70	106	99	117
Estimated median	71	69	68	82	59	61

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

All the coefficients for the variable denoting whether the respondent had read or heard about oil spills (READR) are positive and are statistically significant for the linear 200,000-bird case. The coefficients for the variable denoting whether the respondent agreed that the death of waterfowl is a serious problem (O_DEADD) also are positive and significant for the linear 2,000-bird case. Waterfowl hunters (P_HMWA) have significantly greater WTP for both 2,000-bird specifications and the linear 200,000-bird specification. The coefficient on respondent age (AGEYR) is negative and significant or close to significant for the linear and log 20,000-bird and linear 200,000-bird cases. All four of these variables are retained in the final model used to test for theoretical validity.[3]

Goodness-of-fit statistics for all models and all versions are consistently poor. The best pseudo-R^2 obtained is only 0.14 for the linear 20,000-bird case for the full model. The small R^2s do not invalidate the signs and statistical significance of the estimated coefficients. However, they indicate that conceptually justifiable specifications do not contribute much to explaining variations in WTP among respondents. Although such poor fits could arise from inappropriate functional forms, we generally have been unable to improve fits with more general functional forms. The poor fits may indicate that the process that generates expressed nonuse values is different than that posited by conventional utility theory.

Tables 5-4 through 5-6 also provide estimates of mean and median WTP for each model. The differences among estimated mean WTP for the constant-only, full, and final models are small for the linear case. Differences in means between the linear final and full models are only \$1, \$6, and \$3 for the three treatments, respectively. However, estimated means are more sensitive to model specification for the log case. The estimated means for the final and full models are identical for the log 20,000-bird case, but differ by \$15 and \$18, respectively, for the 2,000-bird and 200,000-bird cases. The differences between the log constant-only and final models are smaller. In any case, none of these differences is large enough to affect the statistical tests of the effects of treatment reported below for the final models.[4]

[3] The criterion for selection in the final model was that the variable coefficient was significant in at least two linear specifications in the full model.

[4] The exclusion of variables with some significant coefficients (MALL, K_COUNT, and INCOME) in the final models also has no material effect on estimated means.

All treatments were carefully randomized between and within mall samples. Nevertheless, any comparisons between treatments are conditional on the characteristics of the subsample to which the treatment was administered. Thus it is important to determine to what extent observed differences among treatments are due to random differences in subsamples rather than the treatment themselves. Tests of differences in variable distributions among subsamples indicate a few minor differences. To determine whether these differences influence our results, we calculated means by combining the estimated coefficient from one treatment with the sample means for those characteristics from other treatments. The largest difference due to sample characteristics is $4. These differences are comparable to those induced by differences in model specification and are too small to affect the results of our evaluation of differences in treatment.[5]

In order to test for differences in mean WTP across treatments, it is necessary to derive standard errors by use of a bootstrap procedure. Our study is one of the first to evaluate differences in WTP estimates with appropriate statistical tests. Before the introduction of the bootstrap to environmental economics by Park, Loomis, and Creel (1991), error bounds were rarely reported.[6] Bootstrapping is a resampling approach to estimating a distribution. The technique presumes that the original sample truly is representative of the population from which it was drawn. Sampling with replacement from the original sample thus simulates variability that would occur if new samples were drawn from the population itself.[7] That is, the routine randomly selects an observation for the bootstrap sample and replaces it so it is eligible for selection again. This selection and replacement process continues

---

[5] Calculations for other models and treatments produced similar results. For example, the differences among linear open-ended spill treatments were less than $1. Differences for log models were larger, but still not large enough to affect difference-of-means tests.

[6] For example, in the absence of a variance estimate, Seller, Stoll, and Chavas (1985) calculate 95-percent use-value confidence intervals from the upper- and lower-bound values of their estimated visit coefficients. This approach ignores variation in the other coefficients and variation that would occur in repeated samples from the same population.

[7] Both Park, Loomis, and Creel (1991) and Hanemann, Loomis, and Kanninen (1991) bootstrap the WTP variance by simulating the multivariate normal distribution of the coefficients rather than by resampling the actual data. This approach assumes that the original set of coefficient estimates is the "truth." Although resampling is cumbersome and time-consuming because it involves repeatedly re-estimating the coefficients themselves, it is preferable because it allows variation in repeated samples from the same population to determine the variance.

until the bootstrap sample is the same size as the original data set. However, each bootstrap sample has a different composition. One observation may be repeated, while another observation may not appear at all. We then re-estimate the model for the resampled data and calculate the associated mean WTP. Our models were bootstrapped using 1,000 iterations. The standard deviation of the distribution of those 1,000 means is then the standard error of the mean of the original data.

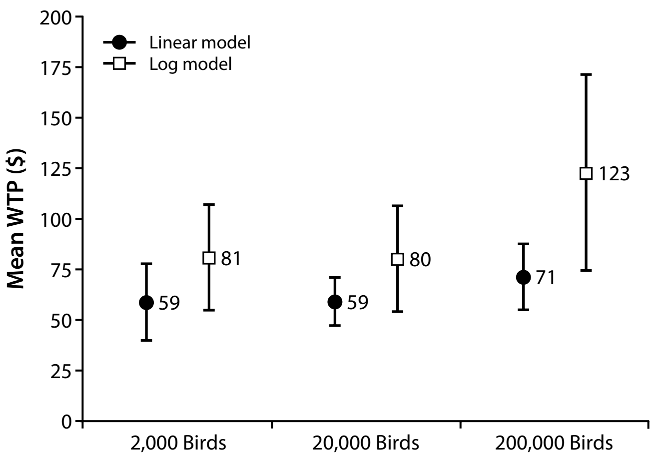
Figure 5-2 shows 90-percent confidence intervals calculated with the resulting bootstrapped means and standard errors. The linear model for the 2,000-birds version and for the 20,000-birds version have the same mean WTP, while the mean WTP for the 200,000-birds version is somewhat higher. The means for the log models are all higher than their linear counterparts. Again the 2,000-birds version and the 20,000-birds version have virtually the same WTP, while the third version is higher. Although the means increase as the level of protection increases, the confidence intervals substantially overlap for both the linear and the log case, implying that the means may not be significantly different.

Z-statistics, shown at the bottom of Table 5-7, test for differences across versions and serve as our test of theoretical validity. The results reinforce the visual evidence from Figure 5-2. Four of six cases, including all three based on the linear model, show no statistically significant differences in predicted mean WTP. The comparisons between the log-model means for 200,000 birds and the other two versions are marginally significant at the 10-percent level. However, Tables 5-4, 5-5, and 5-6 indicate that the linear models have better fits and more significant coefficients. Moreover, the small estimated values of cc reported in Appendix F for the general Box-Cox model are more consistent with the linear specification than the log specification. Therefore, we fail to reject the null hypothesis of no difference in the means across questionnaire versions. This result is inconsistent with theoretical validity.[8]

[8] The effective sample size shown in Table 5-7 is about 400 for each version, after cleaning the data. If we hold the mean and variance constant and increase the sample size, the increase in N would decrease the standard error and increase the calculated z- values. Under these assumptions, the observed differences between the 200,000-bird version and the other two versions for the linear model would be significant if the sample sizes increased to about 1,000 and 700 for the 2,000-bird and 20,000-bird versions, respectively. In effect, the sample sizes would have to at least double before the observed differences would be statistically significant.

We also performed a likelihood-ratio test to compare the vectors of coefficients from each of the migratory-waterfowl models run independently with a model run on the pooled data sets. This test statistic, which is distributed chi-square, indicates whether the signs and sizes of the coefficients vary significantly across treatments. The test statistics for the migratory-waterfowl models are 50.6 for the linear models and 152.9 for the log models. Both of these statistics are significant at the 1-percent level, allowing us to reject the null hypothesis that the vectors of coefficients are the same across versions. This finding is not surprising in view of our discussion of the coefficients of the models. However, it is somewhat contrary to the findings of the nonparametric tests and means tests, which show no significant difference in the distributions or in the means of WTP. The significance of the likelihood-ratio test implies that, although we observe identical distributions and WTP means, the processes generating these WTP values are different.[9]

Figure 5-2. Predicted Means and 90-Percent Confidence Intervals for Migratory-Waterfowl Data



[9] This result is not an artifact of the relatively poor fits for all the models. Rather, it indicates basic structural differences in the proportion of the variation that is explained.

Our results for the migratory-waterfowl tests show that stated WTP values are statistically constant regardless of whether 2,000, 20,000, or 200,000 bird deaths are prevented. This result occurs for nonparametric and parametric tests of the hypothesis. One exception comes from the likelihood-ratio test, which shows that the vectors of estimated coefficients are not identical across treatments. However, the remainder of our tests are consistent. Given the basic theoretical assumption that the marginal utility of preventing additional deaths is positive, our results indicate that theoretical validity of the estimates does not hold in this experiment.

Table 5-7. Z-tests of Differences in Predicted Mean WTP: 2,000 Birds, 20,000 Birds, and 200,000 Birds

	Linear Model	Log Model
2,000 Birds		
Mean	59	81
Standard error	11	16
N	267	267
20,000 Birds		
Mean	59	80
Standard error	7	16
N	271	271
200,000 Birds		
Mean	71	123
Standard error	10	29
N	264	264
z-statistics for Theoretical Validity Test (One-Tailed Test)		
2,000 Birds < 20,000 Birds	-0.02	0.06
20,000 Birds < 200,000 Birds	-0.98	-1.30*
2,000 Birds < 200,000 Birds	-0.82	-1.26*

* Significant at the 10% level for a one-tailed test.
** Significant at the 5% level for a one-tailed test.
*** Significant at the 1% level for a one-tailed test.

5.3 Oil Spills (Open-Ended)

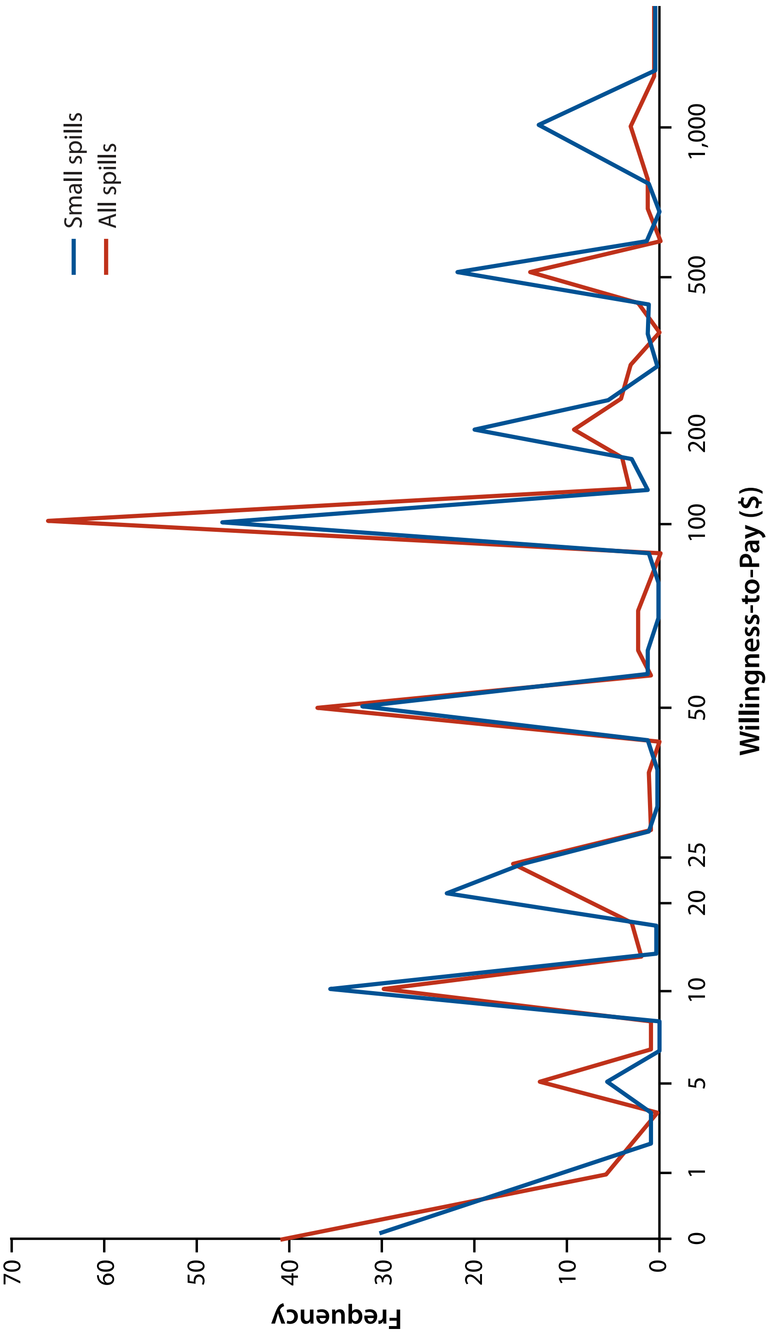
The second test of theoretical validity is whether there are any differences in WTP responses for reducing the environmental effects from only small oil spills compared to reducing the environmental effects from small and large (i.e., all) spills, using open-ended data. As before, we begin by presenting the univariate statistics for WTP by version, shown in Table 5-8. A surprising pattern is revealed: the mean WTP for the all-spills version ($95) is substantially less than the mean for small spills ($144). However, as we saw with the migratory-waterfowl data, the medians and ranges are consistent across versions, and the standard deviations are quite large. The graphs of the WTP responses are more informative. As shown in Figure 5-3, the distributions are very similar. The peaks and valleys are quite consistent across the entire range.

Table 5-8. Univariate Results for Two Open-Ended Oil-Spills Questionnaires: Censored Data

	Censored Data	
	Small Spills (WTP $)	All Spills (WTP $)
Mean	144	95
Standard Deviation	239	157
Median	50	50
Mode	100	100
Range	0–1,000	0–1,000
Shapiro-Wilk statistic[a]	0.60	0.59
N	275	282

[a] This statistic indicates that these distributions are not normal.

Figure 5-3. WTP Frequency Distribution: Oil Spills



To formally evaluate the similarity of the distributions, we again use the nonparametric WRS and permutations tests (see Table 5-9). The p-values shown in the table (0.917 and 0.998) reflect the probability that WTP for small spills is less than WTP for all spills. One minus this value would be the probability that WTP for small spills exceeds WTP for all spills. These probabilities (0.083 and 0.002) are statistically significant. These nonparametric results indicate the same unexpected relationship between WTP and the level of nonuse services that we saw with the univariate statistics. The two WTP distributions are significantly different, but the small-spills distribution is substantially higher than the all-spills distribution.

As an additional test of theoretical validity, we use a z-test to analyze the significance of differences in mean WTP for the small-spills and all-spills versions of the oil-spills questionnaire, as estimated by tobit models. Table 5-10 presents a list of variables included in the models. Tables 5-11 and 5-12 present the estimated tobit equations for the small-spills and all-spills versions, respectively, using two different functional forms.

As before, we report constant-only, full, and final specifications for each functional form and treatment in Table 5-11 and Table 5-12. The qualitative results for the full models are very similar to the migratory-waterfowl results. Estimates generally have expected signs, but very few are statistically significant. Unexpected negative signs on R_ENV12, P_OILD, and SIZE_MIN and unexpected positive signs on O_NEGSA are not significant. Fits are very poor except for the linear all-spills case, which has a pseudo-R^2 of 0.21.

Table 5-9. Nonparametric Tests of Differences in Open-Ended WTP Distributions: Small Spills and All Spills[a]

Version Comparison	Monte Carlo p-value Estimates[b]
Small Spills vs. All Spills	
Wilcoxon Rank Sum Test	0.917
Permutations Test	0.998

[a] Test statistics are Monte Carlo p-value estimates.
[b] These p-values show that the WTP for small spills significantly exceeds the WTP for all spills.

Table 5-10. Description of Explanatory Variables in Tobit Analysis of Open-Ended Oil-Spills Questionnaires and Probit Analysis of Dichotomous-Choice Oil-Spills Questionnaires

Variable	Description
MALL	Dummy = 1 for Southlake Mall (0 for Lakeshore Mall)
R_ENV12	Dummy = 1 for respondents who ranked environmental pollution first or second in seriousness among 6 social and economic problems (Q.1)
P_OILD	Dummy = 1 for respondents who placed a high priority on government funding for addressing oil spills from tankers (Q.2)
READR	Number of times the respondent had read or heard about U.S. oil spills in the 6 months prior to the survey (Q.3)
OPA	Dummy = 1 for respondents who had heard or read about the Oil Pollution Act (Q.4)
S3_CNT4	Number of oil spills (excluding the *Exxon Valdez* oil spill) rated as extremely serious (Q.9)
SIZE_MIN	Dummy = 1 for respondents who agree or strongly agree that spills of any size cause serious environmental damage *and* disagree or strongly disagree that some very large spills cause only minimal environmental damage (Q.11)
O_NEGSA	Dummy = 1 for respondents who strongly agree that most oil spills are the result of oil company negligence (Q.11)
O_EXIST	Dummy = 1 for respondents who disagree or strongly disagree that people have a right to change the environment to meet their needs, ***and*** who agree or strongly agree that the environment should be preserved at all cost (Q.10)
AGEYR	Age of respondent (midpoint of categories in Q.13, with 75 for the highest category)
NORGS	Number of environmental or conservation organization memberships in the respondent's household (Q.19)
INCOME	Respondent's household income (Q.14)

Note: Specific questions are referenced following the description. For example, (Q.8) refers to question 8 of the survey.

Table 5-11. WTP Models Based on Tobit Analysis: Small Spills Version (t-statistics in parentheses)

	Linear Model			Log Model		
	Constant Only Model	Full Model	Final Model	Constant Only Model	Full Model	Final Model
INTERCEPT	127.57*** (7.81)	185.74*** (2.76)	8030*** (2.81)	0.005** (2.34)	0.014** (1.98)	0.003 (0.88)
READR		2.79 (0.58)	3.35 (0.71)		0.000 (0.08)	0.000 (0.04)
O_EXIST		66.31 (1.75)*	70.11** (1.99)		0.004 (1.04)	0.004 (1.51)
NORGS		12.46 (0.90)	20.37 (1.50)		0.002 (1.42)	0.001 (1.00)
MALL		-3.32 (0.09)			-0.000 (0.08)	
R_ENV12		14.64 (0.41)			-0.001 (0.30)	
P_OILD		-33.67 (1.01)			-0.001 (0.22)	
OPA		-25.97 (0.79)			0.000 (0.04)	
S3_CNT4		15.72 (0.76)			-0.000 (0.05)	
SIZE_MIN		-8.22 (0.23)			0.001 (0.33)	
O_NEGSA		21.99 (0.52)			-0.000 (0.05)	
AGEYR		-4.02*** (2.91)			-0.000 (1.07)	
INCOME		0.002*** (2.72)			-0.000 (1.60)	
SIGMA	261.92*** (21.49)	250.38*** (21.35)	259.30*** (21.43)	0.017*** (53.39)	0.016*** (29.16)	0.017*** (50.77)
N	266	259	263	266	259	263
Pseudo-R^2	0.01	0.06	0.03	0.01	0.01	0.02
Estimated mean	128	134	129	185	150	196
Estimated median	128	125	111	136	126	131

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 5-12. WTP Models Based on Tobit Analysis: All Spills Version (t-statistics in parentheses)

	Linear Model			Log Model		
	Constant Only Model	Full Model	Final Model	Constant Only Model	Full Model	Final Model
INTERCEPT	77.42*** (7.26)	-65.74 (1.38)	-5.98 (0.31)	0.002*** (3.62)	0.003 (1.03)	0.001 (0.51)
READR		11.32*** (3.86)	11.54*** (3.81)		0.000 (1.42)	0.000 (1.32)
O_EXIST		7.62 (0.32)	41.52* (1.94)		0.001 (0.42)	0.001 (1.29)
NORGS		35.11*** (3.43)	38.02*** (3.71)		0.001 (1.58)	0.001 (1.43)
MALL		-7.77 (0.33)			-0.001 (0.72)	
R_ENV12		11.59 (0.57)			-0.000 (0.33)	
P_OILD		47.23** (2.33)			0.001 (0.72)	
OPA		15.21 (0.76)			0.001 (0.46)	
S3_CNT4		41.51*** (3.49)			0.001*** (2.74)	
SIZE_MIN		51.48** (2.24)			0.001 (0.79)	
O_NEGSA		17.76 (0.61)			-0.000 (0.20)	
AGEYR		-1.18 (1.34)			-0.000 (0.57)	
INCOME		0.001** (2.40)			-0.000 (1.51)	
SIGMA	173.69*** (21.46)	155.00*** (21.08)	164.35*** (21.30)	0.007*** (64.15)	0.007*** (37.33)	0.007*** (58.71)
N	279	265	272	279	265	272
Pseudo-R^2	0.01	0.21	0.14	0.01	0.08	0.04
Estimated mean	77	86	80	93	82	100
Estimated median	77	75	70	100	62	74

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Like the migratory-waterfowl results, the INCOME coefficient is negative for the log specifications, but this time it is (barely) insignificant. Unlike the migratory-waterfowl results, INCOME is positive and significant for both linear cases. A utility-theoretic interpretation is that respondents regard coastal resources threatened by potential oil spills as not being perfectly replaceable (Cook and Graham, 1977).[10] This perception contradicts situations where natural or technical restoration clearly is possible. While differences between perceived and actual risks commonly are observed in other contexts,[11] a competing hypothesis is that nonutility-theoretic factors affect expressed WTP to avoid oil spills. Unfortunately, our data provide no basis for discriminating between these hypotheses.[12]

Turning to the final models, the coefficient for the variable indicating previous information exposure (READR) is significant for the linear all-spills case. The environmental-attitude variable (O_EXIST) and the environmental-organization membership variable (NORGS) are significant in the linear small-spills and all-spills cases, respectively. We use these final models to test for differences in WTP across treatments.

Tables 5-11 and 5-12 also report estimated mean and median WTP for each model. As we saw in the migratory-waterfowl estimate, means based on linear models are relatively insensitive to model specification. The means for the two final models are \$5 and \$6 less than the corresponding full models. However, the log models continue to be much more sensitive to specification. The mean WTP for the small-spills final model is \$46 higher than the full model and \$11 higher than the constant-only specification. The mean WTP for the all-spills final model is \$18 higher than for the full model and \$7 higher than the constant-only specification. Again, our qualitative conclusions from the difference-of-means tests reported below are unaffected by our particular choice of specification.

---

[10] As discussed in Chapter 4, a significant income coefficient implies that the marginal utility of income varies with the level of protection. This would not occur if perfect substitutes were available for the natural resource services.

[11] See, for example, Hogarth and Reder, 1986; Viscusi and O'Connor, 1984; and Slovic, Fischhoff, and Lichtenstein, 1979.

[12] See, however, Schkade and Payne's (1992) parallel study using the migratory-waterfowl questionnaire.

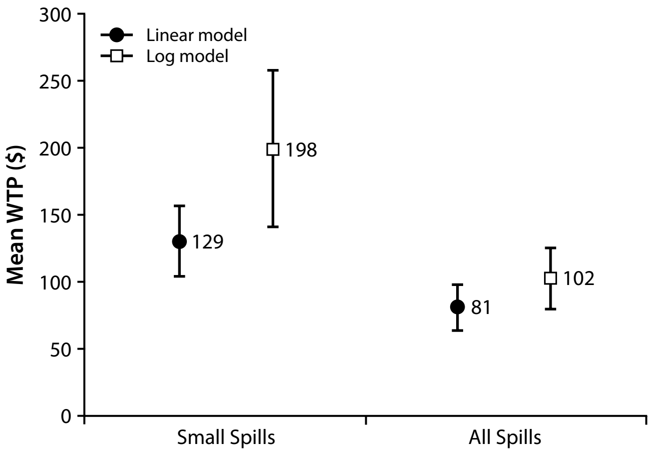
We also evaluate the effect of random differences in treatment subsamples on conditional means. Applying the linear small-spills coefficients to the all-spills sample increases the small-spills mean by $11. Applying the linear all-spills coefficients to the small-spills subsample decreases the all-spills mean by $15. Both of these changes are less than 5 percent of the original estimates, and are not large enough to affect statistical tests of differences in mean WTP for the two oil-spill treatments. Similar results are obtained for the log models.

As in the migratory-waterfowl experiment, we develop bootstrapped means and standard errors for the open-ended oil-spills data from 1,000 iterations of the bootstrap routine. Figure 5-4 shows the resulting mean and standard error for both versions. Once again, the small-spills version has a higher mean and standard error than the all-spills version for both the linear and the log model. Table 5-13 provides the results of the z-tests on the mean WTP from both the small-spills and all-spills versions, using both the linear and log models. The relevant statistical test for theoretical validity is whether the all-spills means are greater than the comparable small-spills means. The z-statistics for this test are shown at the bottom of the table.[13] Although the z-statistics are relatively large for both the linear and log models, they are positive, which implies that they are not statistically significant for the null hypothesis that protecting the environment from all oil spills produces larger values than protecting the environment from only small oil spills.

As with the migratory-waterfowl experiment, we performed a likelihood-ratio test comparing the vectors of coefficients from the two oil-spill models run independently with a model run on the data sets pooled. The test statistics are 59.3 for the linear models and 167.7 for the log models. Both of these test statistics are significant at the 1-percent level, leading us to reject the null hypothesis that the vectors of coefficients are the same across versions. Given the large, albeit counterintuitive, differences in the distributions and mean WTPs, the significance of these test statistics is neither surprising nor contradictory.

[13] As discussed above, the log model yields higher means than the linear model, and this difference is significant for the small-spills version.

Figure 5-4. Predicted Means and 90-Percent Confidence Intervals for Open-Ended Oil-Spill Data



Our results for the open-ended oil-spills experiment show a relationship between WTP values and the level of oil-spill protection that is contrary to expectations. Instead of finding that people are willing to pay more for higher levels of protection, we find that people are willing to pay significantly less. This result occurs for both nonparametric and parametric tests of the hypothesis. As shown in the migratory-waterfowl experiment, assuming positive marginal utility for oil-spill protection, our oil-spill results indicate that theoretical validity of the estimates does not hold.

Table 5-13. Z-tests of Differences in Predicted Mean Open-Ended WTP: Small Spills and All Spills

	Linear Model	Log Model
Small Spills		
Mean	129	198
Standard error	15	35
N	263	263
All Spills		
Mean	81	102
Standard error	9	13
N	272	272
z-statistics for Theoretical Validity Test (One-Tailed Test)		
Small Spills < All Spills	2.68[a]	2.59[a]

* Significant at the 10% level for a one-tailed test.
** Significant at the 5% level for a one-tailed test.
*** Significant at the 1% level for a one-tailed test.

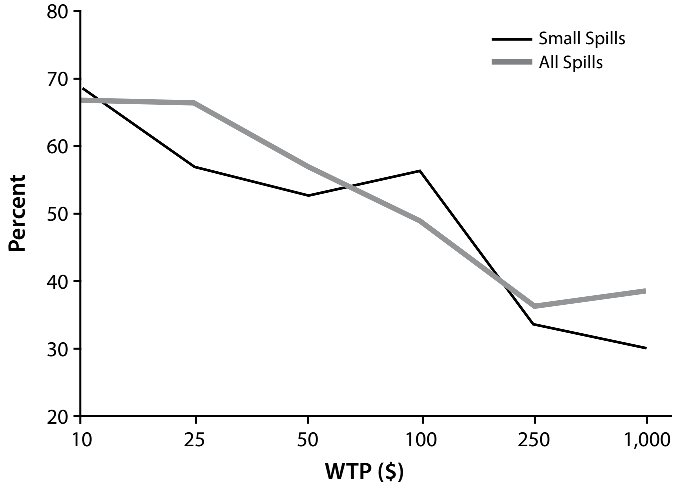
[a] This z-statistic is not statistically significant for the one-sided null hypothesis that the mean of small spills is less than the mean of all spills.

5.4 Oil Spills (Dichotomous-Choice)

The third test of theoretical validity is whether there is a difference in WTP responses for different levels of oil-spill response from the dichotomous-choice CV format. One might argue that the theoretical validity findings from the previous two tests are attributable to the use of the open-ended question format. This section examines the results from our dichotomous-choice survey and their implications for theoretical validity.

Figure 5-5 shows the percentages of respondents who agreed to pay each specified dichotomous-choice bid in the small-spills and all-spills versions. While both graphs generally slope downward as we would expect, each graph has one segment with a positive slope. Furthermore, the all-spills curve does not clearly dominate the small-spills curve at all bid amounts. In order for the all-spills version to yield a higher mean WTP estimate, the area under the all-spills curve would have to be greater than the corresponding area under the small-spills curve.

Figure 5-5. Percentage of Oil-Spill Respondents Agreeing to Pay Selected Amounts in Dichotomous-Choice Format: Small-Spills and All-Spills Versions



To investigate the distributions shown in Figure 5-5, Table 5-14 presents the actual proportions of respondents who accepted and rejected each bid amount for both versions. A chi-square test of the null hypothesis that the distributions of responses to the small-spills and all-spills versions are the same cannot be rejected. In other words, the two distributions are statistically identical, and we cannot conclude that the all-spills data yield a larger measure of central tendency. Theoretical validity, once again, is not established.[14]

As explained in Chapter 4, it is necessary to estimate the cumulative probability of accepting various bid amounts to derive mean values for the dichotomous-choice format. In this analysis maximum-likelihood models are estimated using the linear and log functional forms discussed in Chapter 4. (Table 5-10 describes the variables in these probit models.) The model specifications are very similar to those used

[14] This computation of the chi-square statistic is consistent with a two-tailed test for differences in the distribution. An alternative test, the common odds ratio test, yields the same result for a one-tailed test (i.e., the WTP distribution for all spills is not greater than the WTP distribution for small spills).

Table 5-14. Chi-Square Test for Differences in Distributions of Dichotomous-Choice Responses: Small Spills Versus All Spills Versions

	Small Spills	All Spills
BID = 10		
No	20	22
Yes	44	44
BID = 25		
No	29	22
Yes	38	43
BID = 50		
No	29	28
Yes	32	37
BID = 100		
No	28	35
Yes	36	33
BID = 250		
No	42	41
Yes	21	23
BID = 1,000		
No	45	39
Yes	19	24
	Chi-Square = 3.44[a]	

* Significant at the 10% level for a two-tailed test.
** Significant at the 5% level for a two-tailed test.
*** Significant at the 1% level for a two-tailed test.

[a] This chi-square is not statistically significant at the 10% level.

for the tobit estimates in the previous section (see Tables 5-11 through 5-13).[15] We report three models for each treatment: bid only, full model, and a final model.

As before, coefficients for the full models reported in Tables 5-15 and 5-16 generally have the expected signs but few are statistically significant and the fits are unimpressive. The MALL coefficients are insignificant, indicating no difference in responses between mall samples. INCOME also is insignificant, except for the linear, all-spills case, where it is positive and barely significant. Again, this is weak evidence of possible perceived irreplaceability of coastal resources or possible nonutility-theoretic responses. (See discussion above on open-ended responses.)

Looking at the final models, the coefficients for the variables denoting whether the respondent had read or heard about oil spills (READR), and the number of membership organizations (NORGS) are statistically significant for the small-spills treatment. However, READR is negative, indicating that better-informed respondents are less likely to accept the offered bid. Our a *priori* expectation was that this coefficient should be positive. However, if better informed respondents have more information about recoverability, this negative sign would not be inconsistent. The coefficient for the pro-environmental variable (O_EXIST) is significant for the all-spills treatment. The final models are used to test theoretical validity.

The estimated means generally are insensitive to model specification for the linear functional form. The linear, small-spills final model has a mean of $239 compared with $232 and $231 for the other two specifications. The linear, all-spills final model has a mean of $342 compared with $342 and $353 for the other two specifications. However, the log means are more sensitive to model specification. The small-spills final model has a mean of $262 that is much higher than the means of $192 and $218 for the other two specifications. In contrast, the all-spills final model has a mean of $167 that is much lower than the means of $204 and $305. These differences apparently arise from the somewhat poorer fits for the log specifications compared with the linear specifications. Flexible functional form estimates reported in Appendix F indicate that the linear specifications appear to be closer to the proper form than the log specification.

---

15 SIGMA is the estimate of σ defined in Equations 4.2 and 4.4. In probit models,

$$\sigma = -\frac{1}{b}$$

where b is the coefficient on the bid variable.

Table 5-15. Dichotomous-Choice Models Based on Probit Analysis: Small Spills Version (t-statistics in parentheses)

	Linear Model			Log Model		
	BID Only Model	Full Model	Final Model	BID Only Model	Full Model	Final Model
INTERCEPT	232.01** (2.37)	-150.72 (0.42)	331.96** (2.30)	0.009 (1.36)	-0.014 (0.55)	0.019* (1.72)
SIGMA	1232.02*** (4.20)	1130.66*** (4.34)	1133.77*** (4.42)	0.096*** (3.12)	0.080*** (3.40)	0.088*** (3.34)
READR		-60.80*** (2.58)	-48.69** (2.14)		-0.004*** (2.64)	-0.004** (2.18)
O_EXIST		48.65 (0.27)	45.50 (0.27)		0.004 (0.30)	0.002 (0.18)
NORGS		167.79* (1.91)	204.49** (2.42)		0.011 (1.73)	0.014** (2.16)
MALL		-37.28 (0.21)			-0.001 (0.09)	
R_ENV12		207.86 (1.28)			0.016 (1.42)	
P_OILD		-191.00 (1.15)			-0.012 (1.04)	
OPA		468.60*** (2.87)			0.034*** (2.97)	
S3_CNT4		142.07 (1.42)			0.011 (1.49)	
SIZE_MIN		240.77 (1.39)			0.015 (1.26)	
O_NEGSA		-225.95 (1.14)			-0.014 (1.00)	
AGEYR		4.61 (0.68)			0.000 (0.97)	
INCOME		132.89 (0.36)			-0.017 (0.61)	
N	371	362	365	371	362	365
Chi-Square	18.35	44.27	28.16	11.13	37.72	20.57
% correctly predicted	61.73	62.71	61.64	56.33	62.43	60.27
McFadden R^2	0.04	0.09	0.06	0.02	0.08	0.04
Estimated mean	232	231	239	192	218	262
Estimated median	232	230	239	374	347	415

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Table 5-16. Dichotomous-Choice Models Based on Probit Analysis: All Spills Version (t-statistics in parentheses)

	Linear Model			Log Model		
	BID Only Model	Full Model	Final Model	BID Only Model	Full Model	Final Model
INTERCEPT	342.32*** (2.69)	581.98 (1.16)	124.95 (0.64)	0.022* (1.67)	0.074 (0.95)	-0.004 (0.16)
SIGMA	1622.34*** (3.26)	1509.03*** (3.33)	1500.98*** (3.43)	0.185* (1.72)	0.234 (1.24)	0.189* (1.66)
READR		-16.34 (0.58)	-19.32 (0.72)		-0.002 (0.55)	-0.002 (0.68)
O_EXIST		744.19*** (3.20)	722.50*** (3.42)		0.111*** (3.11)	0.090*** (3.39)
NORGS		73.44 (0.81)	112.94 (1.32)		0.012 (0.84)	0.014 (1.32)
MALL		-325.20 (1.32)			-0.045 (1.18)	
R_ENV12		221.66 (1.03)			0.036 (1.08)	
P_OILD		-218.20 (0.98)			-0.026 (0.77)	
OPA		153.29 (0.72)			0.023 (0.69)	
S3_CNT4		-69.98 (0.53)			-0.012 (0.60)	
SIZE_MIN		262.01 (1.15)			0.041 (1.16)	
O_NEGSA		-193.64 (0.71)			-0.026 (0.63)	
AGEYR		-18.00* (1.76)			-0.003** (1.97)	
INCOME		703.41* (1.66)			0.090 (1.32)	
N	378	370	375	378	370	375
Chi-Square	10.79	36.39	25.72	2.93	26.59	16.38
% correctly predicted	56.35	61.89	65.07	56.35	59.73	61.60
McFadden R^2	0.02	0.07	0.05	0.01	0.05	0.03
Estimated mean	342	353	342	204	305	167
Estimated median	342	353	342	882	1,367	871

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

The final model specifications shown in Tables 5-15 and 5-16 provide the basis for generating bootstrapped means and standard errors for each questionnaire version. To test for theoretical validity, we compare the bootstrapped means for reducing environmental effects from small spills with the corresponding mean for reducing environmental effects from all spills.

Figure 5-6 shows the confidence intervals for each version of the dichotomous-choice oil-spill data, estimated with two alternative functional forms. For each functional form/estimator, the intervals for the two versions overlap substantially.

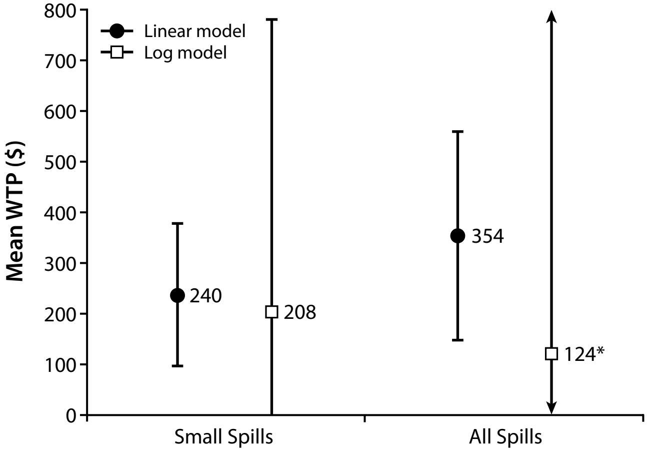
As shown in Table 5-17, the z-statistics indicate that the all-spills means are not statistically different from the comparable small-spills means.[16] These results are consistent with the results of the chi-square tests and the visual evidence from Figure 5-6. They confirm that theoretical validity does not hold for these dichotomous-choice estimates.

The likelihood-ratio test also confirms that there are no differences in these models across versions. The test statistics are 7.4 for the linear model and 8.4 for the log model. Neither of these values is statistically significant at the 10-percent level. For dichotomous-choice responses, this test has a slightly stricter interpretation than for open-ended responses. Given that we are estimating a cumulative density function (see Chapter 4), a failure to reject the null hypothesis that these models are the same implies that not only are the vectors of coefficients the same, but all moments of the distribution are the same. We have established with our z-tests that the means are not statistically different. The likelihood-ratio test shows that all other moments of these distributions are not statistically different, as well.

Our results for the dichotomous-choice oil-spills experiment reinforce the findings from the previous two experiments. Again, we find that the expressed WTP for reducing environmental effects from all oil spills does not exceed the corresponding expressed WTP for reducing the environmental effects from only small spills. This result is robust across chi-square tests on the distribution of Yes and No responses, as well as for the estimated means. These dichotomous-choice results indicate again that theoretical validity of the estimates does not hold.

[16] In order to detect a significant difference between the observed small-spill and all-spill values for the linear model, sample sizes would have to be about three times larger, or greater than 1,000 observations. See footnote 8 in this chapter.

Figure 5-6. Predicted Means and 90-Percent Confidence Intervals for Dichotomous-Choice Oil-Spill Data



*The 90% confidence interval for this estimate extends beyond the graph.

Table 5-17. Z-tests for Differences in Predicted Mean Dichotomous-Choice WTP: Small Spills and All Spills

	Linear Model	Log Model
Small Spills		
Mean	240	208
Standard error	85	351
N	365	365
All Spills		
Mean	354	124
Standard error	126	1,579
N	375	375
z-statistics for Theoretical Validity Test (One-Tailed Test)[a]		
Small Spills < All Spills	-0.75	0.05

* Significant at the 10% level for a one-tailed test.
** Significant at the 5% level for a one-tailed test.
*** Significant at the 1% level for a one-tailed test.

[a] None of these test statistics is statistically significant at the 10% level.

5.5 Summary and Implications

Our results clearly indicate a lack of theoretical validity, because in nearly all cases providing higher levels of environmental protection does not lead to larger estimates of benefits. This finding holds for two very different types of environmental services: protecting migratory waterfowl and preventing damage from oil spills. For the oil-spill experiment, the finding holds for both the open-ended and dichotomous-choice questioning formats. In fact, the open-ended oil-spill data show a counterintuitive result where the small-spills means are significantly larger than the all-spills means. Finally, we employ multiple statistical tests for all three data sets. The results from these tests, taken as a whole, reject theoretical validity of the estimates, calling into question CV's ability to produce valid estimates of the value of natural resource services that consist largely of nonuse values.

CHAPTER 6

Tests of Convergent Validity

6.1 Background

Researchers have used four formats to ask CV questions: iterative bidding, open-ended, dichotomous-choice, and payment cards. Davis (1964) and Randall, Ives, and Eastman (1974) used iterative bidding in their seminal CV studies, but questions regarding the effects of starting bids on final valuation responses, starting point bias, resulted in a substantially diminished use of this format (Boyle, Bishop, and Welsh, 1988; and Smith, Desvousges, and Fisher, 1986). Payment cards, proposed by Mitchell and Carson (1981), have never experienced widespread acceptance among CV practitioners. These exclusions leave open-ended and dichotomous-choice as the two CV questioning formats employed most frequently in published studies (Mitchell and Carson, 1989).

First used by Bishop and Heberlein (1979) (see also Bishop, Heberlein, and Kealy, 1983), dichotomous-choice has become the most popular questioning format for eliciting WTP values. Hanemann (1984) and Hoehn and Randall (1987) have contributed to the theoretical appeal of dichotomous-choice, and Bishop and Heberlein (1990) have demonstrated criterion validity of dichotomous-choice estimates of use values (see also Heberlein and Bishop, 1986). Accompanying these theoretical and empirical contributions, a number of intuitive arguments have been presented favoring dichotomous-choice questions over open-ended questions:

- It places less burden on respondents, thus increasing response rates.
- It is more similar to choice problems encountered in market and voting settings.

- There is diminished opportunity for strategic behavior.
- It is less likely that responses will be influenced by various experimental or survey-design effects.
- Valuation responses can be modeled in a utility-theoretic context.

To our knowledge, no theoretical or empirical contributions in the literature clearly establish or refute the first three arguments. The fourth argument loosely follows from the theoretical work of Hoehn and Randall (1987) on incentive compatibility, but no empirical work has been conducted to determine whether this assertion holds. The final argument, as we have demonstrated in Chapter 4, is not unique to dichotomous choice. In fact, open-ended and dichotomous-choice questions can, at least at a conceptual level, be modeled to estimate the same theoretical construct.

The dichotomous-choice format greatly reduces the amount of information received from each respondent. Rather than stating a specific dollar amount, respondents merely accept or reject one of a few randomly assigned bids. A dichotomous-choice sample size, therefore, must be larger than open-ended samples to derive comparably precise mean-value estimates.[1]

Our appraisal of the existing literature suggests that the question of whether dichotomous-choice questions are superior to open-ended questions is far from being resolved. In an attempt to sort out this confusing debate, we evaluate the convergent validity of responses to the open-ended and dichotomous-choice question formats from the oil-spill experiment. CV estimates are valid in this sense when two elicitation formats yield comparable values. Testing the convergent validity of CV estimates for the oil-spill experiment involves investigating whether respondents' answers to open-ended and dichotomous-choice questions provide the same estimates of WTP. This test can be accomplished by asking whether the estimated WTP distributions are statistically identical. Furthermore, the

[1] In addition to finding much larger variation in dichotomous-choice estimates, Cameron and Huppert (1991) also detect an apparent upward bias in the WTP estimates. Alberini (1991a) shows that poorly designed dichotomous-choice surveys may require samples that are orders of magnitude larger than open-ended surveys to obtain similarly efficient estimates. Of course, the higher response rate in dichotomous-choice surveys could offset this effect slightly.

introduction of bootstrapping techniques to the nonmarket valuation literature makes it possible to statistically compare estimated means from open-ended and dichotomous-choice questions (Park, Loomis, and Creel, 1991). The results of these tests are crucial since mean estimates of individual damage are used most often to develop aggregate damages for NRDAs.

In most of the previous evaluations of the convergent validity of question formats, participants typically have answered questions in both formats in the same survey instrument (Boyle and Bishop, 1988; Kealy, Dovidio, and Rockel, 1988; Loomis, 1990; and Stevens et al., 1991). Only two published studies have employed independent samples to compare responses to open-ended and dichotomous-choice questions (Johnson, Bregenzer, and Shelby, 1990; Sellar, Stoll, and Chavas, 1985). Unfortunately, these studies do not use bootstrapping techniques for computing standard errors of dichotomous-choice means. Therefore, they can only draw qualitative conclusions when comparing responses to the two questioning formats. Our experimental design uses randomly stratified subsamples, requiring each respondent to answer only one valuation question. This design provides the first unequivocal comparison of the open-ended and dichotomous-choice formats, using appropriate statistical tests, allowing clear inferences regarding convergent validity to be drawn.

In addition to directly comparing the open-ended and dichotomous-choice estimates, we conduct an additional test. In principle, if open-ended and dichotomous-choice questions measure the same theoretical construct, then it should be possible to apply dichotomous-choice bid amounts experimentally to open-ended responses and simulate what respondents' answers would have been if they had been asked a dichotomous-choice question.[2] We then compare these "synthetic" dichotomous-choice distributions with the corresponding actual dichotomous-choice distributions. We also estimate mean WTP from the synthetic data to compare the results with estimates from the actual dichotomous-choice surveys.

[2] This test is in the spirit of the Cameron and Huppert (1991) experiments comparing use value estimates from a synthetic dichotomous-choice data set derived from a payment-card survey.

6.2 Comparing Dichotomous-Choice and Open-Ended Estimates

To conduct the first test of contingent validity between the dichotomous-choice and open-ended question formats, we compare open-ended means from the linear and log models with the corresponding dichotomous-choice means. To make the comparisons consistent, we use the same explanatory variables in each model. This strategy reduces one source of variation that could confound the comparison between the alternative formats. All estimated means and standard errors are derived via bootstrapping. This 2 × 2 design results in the four statistical comparisons shown in Table 6-1. The table contains summary statistics and z-statistics for the test of the null hypothesis that the open-ended mean is not statistically different from the comparable dichotomous-choice mean.

In all four of the comparisons, the dichotomous-choice means and standard errors are larger than the comparable open-ended means and standard errors, a direct result of the thick tail of the dichotomous-choice distribution. However, only one of the four statistical tests results in a significant difference. For the all-spills version, the linear dichotomous-choice mean of $354 is significantly different from the open-ended mean of $81. The inability to reject the null hypothesis for the small-spills, linear model appears to be a result of the large standard error for the dichotomous-choice estimate. The differences in the means for the log models are much smaller, but the differences in the standard errors are much larger. However, as discussed in Chapter 5, we favor the linear model. These mixed statistical results neither refute nor conclusively establish convergent validity of the open-ended and dichotomous-choice estimates.[3]

[3] A recent adaption of the dichotomous-choice format is a double-bounded approach (Hanemann, Loomis, and Kanninen, 1991). The original dichotomous-choice framework involves offering a dollar bid, for a specified level of environmental quality, to which respondents either answer Yes or No. Hanemann and associates refer to this as the single-bounded approach. The double-bounded approach involves a second iteration. Respondents who answer Yes to the first bid are presented with a second, higher bid to either reject or accept. Conversely, for those who reject the initial bid, a second, lower bid is suggested. Hanemann, Loomis, and Kanninen show that estimated double-bounded means are, on average, 21 percent less than traditional single-bounded, dichotomous-choice means. In addition, 90-percent confidence intervals around the double-bounded means are an average of 83 percent smaller than the 90 percent confidence intervals for the comparable single bounded means. McFadden and Leonard (1992) obtain similar results.

Table 6-1. Z-tests for Differences in Predicted Mean WTP: Open-Ended and Dichotomous-Choice Data Sets[a]

Model	Open-Ended Data Set		Dichotomous-Choice Data Set
Linear Model			
Small Spills			
Mean	129		240
Standard error	15		85
N	263		365
z-statistic for test of convergent validity		1.29	
All Spills			
Mean	81		354
Standard error	9		126
N	263		375
z-statistic for test of convergent validity		2.17**	
Log Model			
Small Spills			
Mean	198		208
Standard error	35		351
N	272		365
z-statistic for test of convergent validity		0.03	
All Spills			
Mean	102		124
Standard error	13		1,579
N	272		375
z-statistic for test of convergent validity		0.01	

* Significant at the 10% level for a two-tailed test.
** Significant at the 5% level for a two-tailed test.
*** Significant at the 1% level for a two-tailed test.

[a] Means and standard errors are derived from 1,000 iterations of the bootstrap routine.

6.3 Comparing Dichotomous-Choice Estimates and Synthetic Dichotomous-Choice Estimates

We further explore convergent validity of the two question formats by constructing a synthetic dichotomous-choice data set from the open-ended data. To construct this data set, we randomly assigned one of the six dichotomous-choice bids to each open-ended respondent. This assignment of bids mirrors the random assignment of bids in the dichotomous-choice survey design. When the randomly assigned bid exceeded the respondent's open-ended response, we assumed that the respondent would have answered No. Likewise, when the randomly assigned bid was equal to or lower than the respondent's open-ended response, we assumed that the respondent would have answered Yes. In this manner we created synthetic dichotomous-choice data sets from the responses to the open-ended questions for both the small-spills and the all-spills versions. We then analyzed these synthetic data as if they were actual dichotomous-choice responses. The construction of the synthetic dichotomous-choice data provides more information for comparing open-ended and dichotomous-choice questions than the simple, direct test. Instead of comparing only the means, we can compare the distributions of Yes responses.

We present two types of statistical tests for comparing the synthetic dichotomous-choice data with the actual dichotomous-choice data. The first procedure compares the percentage of respondents answering Yes to each of the offered bids across the two data sets. Table 6-2 presents these percentages for the small-spills data and the resulting chi-square test of the null hypothesis that the two distributions of responses are the same. The null hypothesis of no difference in the observed distributions is clearly rejected. Similarly, Table 6-3 presents the percentages for the all-spills data. Again we reject the null hypothesis of no difference.

The reason for rejecting these null hypotheses becomes clear when we graph the percentages of respondents answering Yes for the actual and synthetic dichotomous-choice data sets. Consider the comparison for the small-spills version shown in Figure 6-1. The percentages of respondents accepting a dichotomous-choice bid of $10 is nearly 70 percent for both the synthetic and actual dichotomous-choice data (67 percent and 69 percent, respectively). However, at the other end of the distributions, only 5 percent of the respondents in the synthetic data set would have accepted a bid of $1,000, while 30 percent of the respondents in the actual dichotomous-choice experiment accepted a bid of $1,000. This same pattern holds for the large-spills data (Figure 6-2). The stark difference in the masses in the upper tails of these empirical distributions leads to the rejection of the null hypothesis of no difference in the distributions.

Before administering the survey, we expected the proportion of respondents accepting a bid of $1,000 to be quite low, based on the open-ended pretest used to design the dichotomous-choice experiment.[4] Therefore, it is not surprising that the synthetic dichotomous-choice data conforms better to our prior expectations than the actual dichotomous-choice data. Some researchers have found similar evidence of "starting-point bias" in CV studies using dichotomous-choice questions (Mitchell and Carson, 1989; Randall and Farmer, 1992). Figures 6-1 and 6-2 also suggest why only one of the four null hypotheses is rejected when directly comparing the open-ended means with the dichotomous-choice means. The large mass in the tails of the actual dichotomous-choice distributions results in all of the dichotomous-choice means exceeding the comparable open-ended means. This large mass in the upper tails also inflates the standard errors, thus reducing the ability of the z-tests to identify statistical differences.

[4] Chapter 7 provides further details on our designing of the bid structure for the dichotomous-choice survey instruments.

Table 6-2. Chi-Square Test for Differences in Synthetic and Actual Dichotomous-Choice Distributions: Small-Spills Version

	Synthetic Distribution[a] (# of respondents)	Actual Distribution (# of respondents)
BID = 10		
No	19	20
Yes	39	44
BID = 25		
No	28	29
Yes	28	38
BID = 50		
No	23	29
Yes	35	32
BID = 100		
No	33	28
Yes	17	36
BID = 250		
No	56	42
Yes	5	21
BID = 1,000		
No	54	45
Yes	3	19
	Chi-Square = 30.83***	

* Significant at the 10% level for a two-tailed test.
** Significant at the 5% level for a two-tailed test.
*** Significant at the 1% level for a two-tailed test.

[a] See text for details on the creation of this data set.

Table 6-3. Chi-Square Test for Differences in Synthetic and Actual Dichotomous-Choice Distributions: All Spills Version

	Synthetic Distribution[a] (# of respondents)	Actual Distribution (# of respondents)
BID = 10		
No	18	22
Yes	38	44
BID = 25		
No	31	22
Yes	29	43
BID = 50		
No	26	28
Yes	33	37
BID = 100		
No	34	35
Yes	21	33
BID = 250		
No	51	41
Yes	7	23
BID = 1,000		
No	57	39
Yes	1	24
	Chi-Square = 39.13***	

* Significant at the 10% level for a two-tailed test.
** Significant at the 5% level for a two-tailed test.
*** Significant at the 1% level for a two-tailed test.

[a] See text for details on the creation of this data set.

Figure 6-1. Percentage of Oil-Spill Respondents Agreeing to Pay Selected Amounts in the Actual and Synthetic Dichotomous-Choice Data Sets: Small-Spills Version

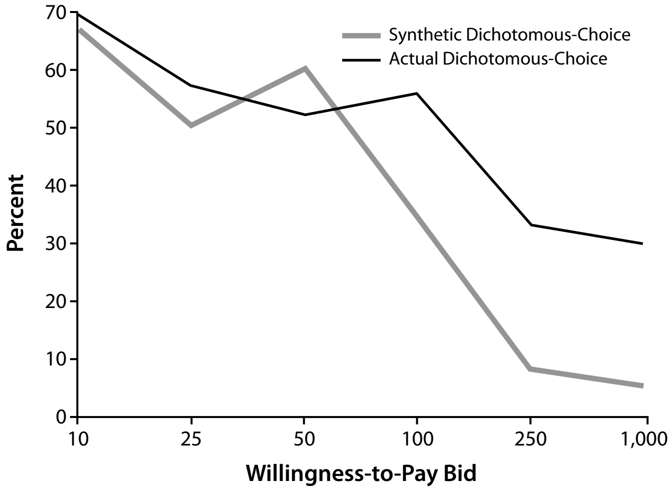
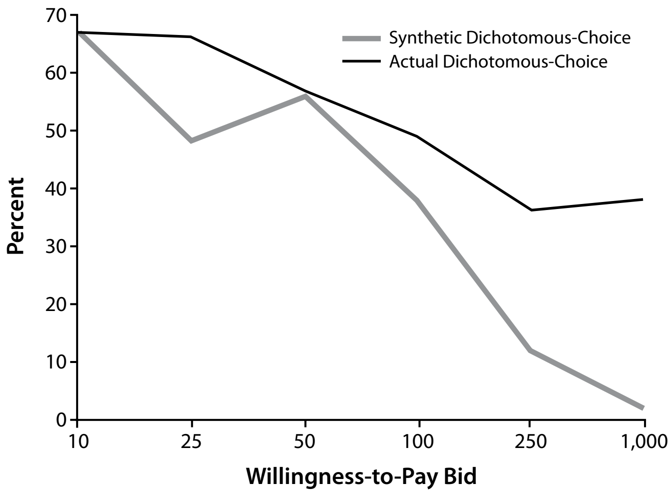


Figure 6-2. Percentage of Oil-Spill Respondents Agreeing to Pay Selected Amounts in the Actual and Synthetic Dichotomous-Choice Data Sets: All-Spills Version



We also can see how the synthetic data behave when estimated within a dichotomous-choice framework, and compare these means with the actual dichotomous-choice means. Table 6-4 presents the results of this difference-of-means test for the small-spills and all-spills data, respectively, and for both the linear and log functional forms. As in the previous comparisons, we derive the means and standard errors using 1,000 bootstrap iterations.

In all four pair-wise comparisons, the actual dichotomous-choice estimates are substantially larger than comparable synthetic estimates. Johnson, Bregenzer, and Shelby (1990) also observe this relationship between their point estimates, with the dichotomous-choice mean nearly two times larger than the open-ended mean. These relationships are not surprising given the differences shown in Figures 6-1 and 6-2. Despite these visual differences, only two of the four statistical tests result in rejection of the null hypotheses of no difference in the estimated means. The two exceptions are the comparisons using the log model estimates for both the small-spills and all-spills versions.[5]

As noted in Chapter 4, we described a Box-Cox flexible functional form. Within the body of this monograph we present two cases that represent the polar special cases of the Box-Cox functional form, the linear and log models. As shown in Appendix F, however, estimation of the Box-Cox function suggests that the linear model fits our data much better than the log model. In turn, the poor statistical fit of the log model and resulting large standard errors, is the likely reason why we cannot reject the null hypotheses of no difference in estimated means for the log model. We, therefore, attach more empirical credence to the results of the linear model than we do to the results of the log model.

[5] Since both the synthetic and actual dichotomous-choice estimates are developed from a single-bound question, we do not need to worry about inefficiencies from not using the double-bounded format proposed by Hanemann, Loomis, and Kanninen (1991). This issue is more of a concern when comparing the actual open-ended means to the actual dichotomous-choice means in Table 6-1.

Table 6-4. Z-tests for Differences in Predicted Mean WTP: Synthetic and Actual Dichotomous-Choice Data Sets[a]

Functional Form and Spill Version	Synthetic Dichotomous-Choice Data Set		Actual Dichotomous-Choice Data Set
Linear Model			
Small Spills Version			
Mean	28		240
Standard error	35		85
N	326		365
z-statistic for test of convergent validity		2.29**	
All Spills Version			
Mean	40		354
Standard error	23		126
N	332		375
z-statistic for test of convergent validity		2.46**	
Log Model			
Small Spills Version			
Mean	-23[b]		208
Standard error	96		351
N	326		375
z-statistic for test of convergent validity		0.64	
All Spills Version			
Mean	81		124
Standard error	35		1,579
N	332		375
z-statistic for test of convergent validity		0.03	

* Significant at the 10% level for a two-tailed test.

** Significant at the 5% level for a two-tailed test.

*** Significant at the 1% level for a two-tailed test.

[a] Means and standard errors are derived from 1,000 iterations of the bootstrap routine.

[b] Negative means are possible with the Hanemann estimator. See Johansson, Kriström, and Maier (1989) for more details.

6.4 Summary and Implications

To examine convergent validity of the open-ended and dichotomous-choice estimates for the oil-spill experiment, we conduct three types of statistical comparisons. The first test directly compares the open-ended and dichotomous-choice means, and the null hypothesis of no difference can only be rejected for one of the four comparisons. However, the failure to reject the other three null hypotheses is a result of large standard errors for the dichotomous-choice means.

The results of the second comparison confirm the above suspicion. For this comparison we create a synthetic dichotomous-choice data set from the open-ended data. Comparisons of the response distributions for the synthetic and actual dichotomous-choice data result in rejection of the null hypothesis of no difference for both the small-spills and all-spills experimental treatments. Rejection of these null hypotheses is a direct result of the large percentages of respondents accepting the two highest bids, $250 and $1,000, in the actual dichotomous-choice experiment. At $1,000 in the synthetic dichotomous-choice data for the small-spills version, the percentages of respondents accepting this bid amount are less than 5 percent, compared to over 30 percent for the actual dichotomous-choice data.

Finally, comparing estimated means from the synthetic dichotomous-choice data with means from the actual dichotomous-choice data results in two of four null hypotheses of no difference being rejected. Specifically, the null hypotheses of no difference in the estimated means are rejected for the linear models, but are not rejected for the log models.

In total, we conduct 10 tests to evaluate convergent validity of the open-ended and dichotomous-choice CV results. For five of these ten comparisons, we reject the null hypotheses of no difference. This overall accounting, however, is somewhat misleading because the failure to reject the null hypothesis occurs primarily (4 of the 5 cases) in the tests using the nonlinear (or log) models, which exhibit substantial variability. Additionally, the log model does not fit the data as well as the linear model (see Appendix F). Because of the poor performance of the log models, we attach more empirical credence to the comparisons conducted using the linear model. For this smaller set of comparisons, the null hypotheses of no difference can be rejected for five of the six comparisons.

As noted by Mitchell and Carson (1989), building on the work of Carmines and Zeller (1979), "convergent validity asks whether [a] measure is correlated with other measures of the same theoretical construct" (p. 191). This is the essence of our comparison of the open-ended and dichotomous-choice estimates for the oil-spill experiment. Both questioning formats are designed to estimate the same theoretical specification of Hicksian surplus. Since it is not known whether either questioning format is valid, Mitchell and Carson (1989) suggest that "[to] the extent that a correlation exists (that is, the measures converge) the validity of each measure is confirmed" (p. 204). Therefore, we reject the null hypotheses of no difference in the open-ended and dichotomous-choice estimates in five of our six tests using our most defensible estimates. We interpret this result as strong evidence for rejecting the proposition of convergent validity between these two commonly used CV questioning formats. Although our data are not sufficient to reveal why respondents answer these two questions differently, Tversky, Slovic, and Kahneman (1990) propose an explanation for possible differences between open-ended and dichotomous-choice measures of WTP. They cite evidence that respondents apply different models of optimizing behavior when confronted with the same decision in a continuous-choice format in contrast with a discrete-choice format, a phenomenon known as "preference reversal."

In the absence of a criterion-validity test we cannot establish which question format provides estimates of value that are closer to the truth. Rejecting convergent validity only indicates that at least one of the question formats provides biased estimates of the welfare change. As noted earlier, Bishop and Heberlein (1990) have conducted a criterion-validity test of dichotomous-choice questions in an experiment employing parallel CV and cash treatments. Since this test was conducted strictly in the context of use values (deer-hunting permits), we believe that it is not appropriate to simply assume that their results apply to nonuse values or to options prices substantially composed on nonuse values. Bishop and Heberlein seem to concur with this conclusion when they state "the overwhelming weight of evidence…favors the use of contingent valuation for estimating willingness to pay for well-defined commodities with private-good characteristics" (p. 101).[6]

[6] This conclusion is reinforced for the criterion-validity experiment conducted by Dickie, Fisher, and Gerking (1987).

Our intuition suggests that the dichotomous-choice format overestimates WTP. It seems implausible that more than 30 percent of our respondents who received the $1,000 dichotomous-choice questionnaire would be willing to write a check for that amount each year to prevent damage from small oil spills, about 3 percent of the average respondent's annual household income. Thus, dichotomous-choice questions appear to induce an anchoring or starting-point bias similar to that observed in early bidding experiments (Boyle, Bishop, and Welsh, 1988; and Smith, Desvousges, and Fisher, 1986).

CHAPTER 7

Dichotomous-Choice Estimation: Tests of Reliability

7.1 Background

The previous chapter discusses various motives for using the dichotomous-choice question format instead of the open-ended format. Despite the widespread popularity of the dichotomous-choice format, there has been little investigation of its accuracy in CV applications. This chapter adds the results of some reliability experiments to our previous analysis of validity, thus completing our assessment of the accuracy of using CV to measure WTP for commodities with substantial nonuse components. As in the previous chapter, we focus our reliability discussion on the oil-spill response part of our experimental design.

Reliability requires that reasonable and justifiable differences in estimation strategy result in statistically comparable estimates. Obviously, model specification and choice of functional form can influence estimates of both open-ended and dichotomous-choice WTP. However, the reduction in respondent burden offered by dichotomous-choice models is accompanied by an increase in estimation burden. Responses contain less information about preferences and thus require more sophisticated estimation techniques. These techniques, in turn, require analysts to make more judgments than are required for open-ended data. For example, analysts must specify a set of bids to offer respondents. The statistical efficiency of these bids depends on guessing correctly about the underlying distribution of WTP (Alberini, 1991a and 1991b; Kanninen, 1990).

The issue is not whether *arbitrary* estimation strategies influence outcomes, but whether a set of plausible strategies that are not clearly superior to each other on theoretical or statistical grounds yield substantially similar WTP estimates. If so, we can judge dichotomous-choice formats to be reliable in this regard. Otherwise, skilled and conscientious analysts could reasonably obtain very different estimates for the same commodity as a result of different, but equally justifiable model choices.

Furthermore, designing a dichotomous-choice survey requires the analyst to specify a particular bid structure when the underlying WTP distribution is unknown. If estimates are robust with respect to reasonable variations in bid structure for the same commodity, then we can judge dichotomous-choice formats to be reliable in that respect.

Finally, estimating a dichotomous-choice model provides a set of coefficients that describes the probability that respondents will accept or reject a given bid. It is necessary to integrate this probability gradient in order to obtain an estimate of mean WTP. Unless the assumed probability distribution restricts nonzero probabilities to the range of plausible WTP values, the analyst must decide which, if any, negative or very large bids will be included in the integral.

This chapter illustrates how the selection among these estimation and experimental design options affects estimated measures of central tendency.

7.2 Choice of Functional Form

The sensitivity of dichotomous-choice CV estimates to the choice of a functional form is an important aspect of reliability. Some recent literature has investigated the effects of functional form (Boyle, 1990; Bowker and Stoll, 1988). Both linear and nonlinear forms are commonly used, some of which are consistent with an underlying utility-theoretic model and some of which are empirical specifications for which there is no corresponding utility function. Proponents of the dichotomous-choice format, such as Hanemann (1984), argue that consistency with utility theory is an important feature of the dichotomous-choice approach. Because we agree that CV estimates are relevant as welfare measures only if they are consistent with welfare theory, we have reported only utility-theoretic estimates in this study.

Table 7-1 reports the bootstrapped means and 90-percent confidence intervals for the zero income elasticity (linear) and unitary income elasticity (log) dichotomous-

Table 7-1. Mean WTP and 90-Percent Confidence Intervals for Alternative Functional Forms: Dichotomous-Choice Format

	Linear Model	Log Model
Small Spills		
Mean	240	208
Confidence Interval	(100, 380)	(-368, 785)
All Spills		
Mean	354	124
Confidence Interval	(148, 561)	(-2472, 2721)

choice oil-spills treatments.[1] The linear and log means for the small-spills treatment differ by only $32, but the difference for the all-spills treatment is $230. However, the wide confidence intervals make it impossible to distinguish these differences statistically. The log models have much wider confidence intervals than the linear models, indicating that the log means are reliable only within about plus or minus 200 percent (for small spills) and 1,500 percent (for all spills) so that the confidence-level lower bounds actually are negative. In contrast, the linear means are reliable within about plus or minus 60 percent and 45 percent, for small and all spills, respectively. However, even these error bounds may not be acceptable for damage-assessment purposes. Specifically, for damage assessments, WTP estimates are multiplied by a measure of the relevant market. A difference of 45 percent, multiplied by thousands (or even millions) of people would result in a large difference in the aggregate damage estimate.

The log models clearly are poorer fits for the data than the linear models.[2] Note that this conclusion is based on the bootstrap variances. CV analysts typically have not performed bootstrap procedures, reporting only the statistical significance and other statistics such as provided in Tables 5-17 through 5-18. This information provides no indication of the relative or absolute reliability of dichotomous-choice means estimates.

[1] The means and standard errors are calculated from the final models reported in Tables 5-15 and 5-16.

[2] The linear model assumes zero income elasticity, which is consistent with the statistically insignificant or small estimates of income elasticity reported in Appendix F.

Figure 7-1 helps illustrate why dichotomous-choice means can be highly variable. The figure shows the zero and unitary income elasticity cases discussed above, plus a popular specification using the logarithm of bid. Although we have no reason to think that respondents have negative values for oil-spill protection, fitting either the linear or log forms implies positive probabilities of negative WTP. If these negative values are ignored in calculating the mean, then the probabilities will not sum to one. The nature of the functional forms requires subtracting the integral of P(No) for negative bids from the integral of P(Yes) for positive bids (Johansson, et al., 1989). Note that the P(Yes) gradient is much flatter for the log form than for the linear form. Specification changes that shift this function slightly have larger effects on the integrals used to calculate the mean for the log form than for the linear form, which accounts for the lower reliability of the log form.

Figure 7-1. Probability of Yes Functions for Three Functional Forms Using the All-Spills, Dichotomous-Choice Data

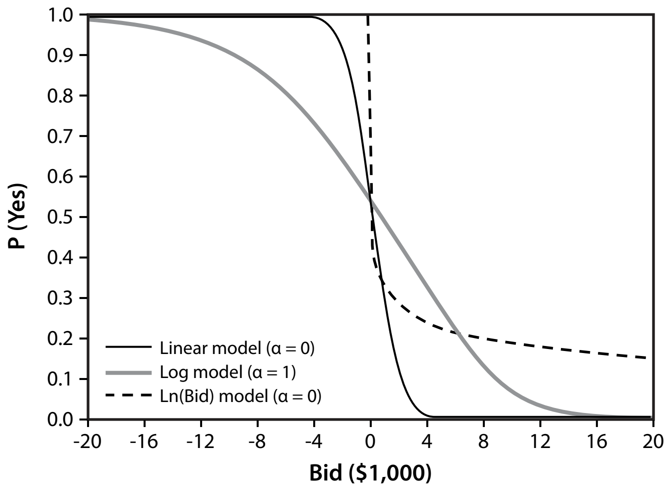


Table 7-2. Effects of Integration-Bound Choice on Estimated Mean WTP for Log(Bid) Model: Small-Spills and All-Spills Versions

Upper Integration Bound	Small-Spills	All-Spills
Income	$4,834	$6,758
25% of Income	1,930	2,428
$5,000	1,217	1,490
$2,000	623	722
$1,000	367	411

The Ln(bid) specification restricts WTP to be positive, but creates an additional problem. The P(Yes) gradient has a very thick tail, which requires choosing an upper bound of integration. Using the final-model specification from Tables 5-17 and 5-18, Table 7-2 compares estimated means for various means for various integration-bound assumptions. Integrating P(Yes) from zero to income for each respondent gives mean WTP of $4,834 and $6,758 for the small-spills and all-spills versions, respectively. Integrating only from zero to $1,000 reduces mean WTP to $367 and $411, respectively. Since there is no conceptual basis for picking an upper bound, different analysts could conceivably obtain WTP estimates ranging between a few hundred dollars and several thousand dollars for the same dichotomous-choice data. The linear and log forms reported in this study have much thinner tails and thus are less sensitive to choice of integration bounds.

From the perspective of consistency among estimates obtained from plausible alternative estimation strategies, the observed high variability in estimates among otherwise reasonable dichotomous-choice specifications indicates serious reliability problems for the dichotomous-choice question format.

7.3 Bid Structure

The second aspect of reliability of dichotomous-choice estimates involves sensitivity to analysts' choice of bid structure. The importance of the dollar value and number of WTP bids randomly assigned to respondents only very recently has begun to receive attention. Cameron and Huppert (1991) estimate the small-sample variation in CV estimates obtained from dichotomous-choice data. They find that the bid structure selected can significantly affect the estimates. Alberini (1991a and 1991b) and Kanninen (1990) have derived the efficiency properties of various bid-structure designs, and conclude that some of the approaches used by CV researchers are seriously flawed. They specifically reject the use of a large number of bids, continuous random bids, and bids that are clustered at low values. All of these approaches yield highly inefficient estimates of desired parameters, thus reducing reliability.

The evaluation of bid structure involves the search for optimal thresholds. Deriving optimal thresholds requires minimizing some value, such as the variance of the estimated mean, that is a function of the number and position of the thresholds. As an example, consider the following simple model:[3]

$$WTP_i = \mu + e_i \tag{7.1}$$

where e_i is normally distributed with $E(e_i) = 0$ and variance equal to one. Assume all respondents are assigned the same bid. The discrete-choice model is a probit equation

$$P(\text{Yes}) = \Phi(\mu - \text{bid}) \tag{7.2}$$

where Φ is the normal cumulative distribution function. The variance of the intercept term μ is

$$\sigma^2 = \frac{\{\Phi(\mu - \text{bid})[1 - \Phi(\mu - \text{bid})]\}}{[n\phi^2(\mu - \text{bid})]} \tag{7.3}$$

where n is the number of observations and ϕ is the normal probability distribution function. The variance σ^2 is minimized when bid = μ, which is the mean and median of a symmetric distribution.

[3] This discussion relies heavily on Alberini (1991a).

Of course, in practice it is desirable to estimate both means and variances and such parameter values are unknown at the survey design stage when the bid structure must be determined. Therefore, generalizations of the theory of optimal experimental design include estimating both location and scale parameters, uncertainty about the parameters of the underlying distribution, and nonnormality. Because the underlying distribution is unknown, researchers may face a tradeoff between efficiency and robustness if their guess about the underlying distribution parameters is wrong.

Although this research has begun to establish systematic procedures for constructing dichotomous-choice thresholds, designing bid structures in practice still requires some guesswork on the part of researchers. Researchers must make assumptions about the functional form, location (mean), and scale (variance) of the underlying WTP distribution and choose the number and values of thresholds appropriate to those assumptions. Alberini (1991a) concludes that a design of between four and eight thresholds that divide the sample into percentiles is a reasonable empirical compromise.

In our experiments, we followed a common procedure to determine the bid structure in the dichotomous-choice versions of the oil-spills survey instrument. We administered an open-ended pre-test and then selected six WTP thresholds based on an analysis of these open-ended responses.[4] We wanted at least 60 Yes/No responses for each bid for statistical tests. Six bids was the maximum number of bids we could have for a total sample size of 400 for each treatment (determined by budget constraints). We selected a low bid ($10) that would be accepted by almost all the respondents for whom preventing environmental damage from oil spills has a positive value.

We tried to select a high bid ($1,000) that would be rejected by almost all the respondents in an effort to determine at what dollar value the probability of Yes function falls to zero. Selecting such a large bid is not the typical practice in dichotomous-choice surveys. However, the pretest indicated that a very large bid

[4] The pretest sample consisted of 77 observations, divided equally between the small-spills and all-spills versions. The pooled versions had a mean of $169 and a median of $100. Nine percent of the stated WTP values were less than $10 and 5 percent were greater than $1,000.

would be needed to define the upper end of the distribution.[5] We selected four intermediate bids ($25, $50, $100, and $250) that approximately divided the pretest sample into percentiles.

Cooper and Loomis (1992) have reported the results of altering the bid structure for estimated deer-hunting use values. Deleting observations corresponding to high bids had the effect of depressing estimated WTP by as much as 57 percent in one case. Table 7-3 illustrates how choice of maximum bid affects results for the linear oil-spill model.[6] We re-estimate the final model using a more typical maximum bid of $250 by deleting the respondents who received the $1,000 bid.[7] Eliminating the highest bid reduces estimated mean WTP by about half for the small-spills treatment, from $212 to $114. The difference is more than three-fold for the all-spills treatment, where the mean falls from $436 to $125. In essence, by removing the $1,000 bid, we allow the probability of Yes function to approach zero at a much lower bid level because there is no information to prevent that occurrence. However, the bootstrap standard error is so large for the comparable all-bids estimates that neither of these differences is statistically significant. Note that the standard error is strongly influenced by the $1,000 bid. The ratio of the standard error to the mean falls from 0.53 to 0.30 for the small-spills version and from 0.55 to 0.27 in the all-spills version when we exclude the $1,000 bid. Clearly the choice of maximum bid and respondents' reaction to that bid has a substantial impact on the reliability of dichotomous-choice estimates.

[5] In fact, the level of acceptances even at $1,000 remained quite high, at 30 percent and 38 percent for small-spills and all-spills versions, respectively. Our attempt to define the upper tail was not particularly successful because of the difference in the shape of the open-ended and dichotomous-choice distributions. As illustrated in Figures 6-1 and 6-2, the dichotomous-choice tails are much thicker than the open-ended tails. Such fat tails are not unusual in nonuse dichotomous-choice surveys. For example, a study of the Kakadu Conservation Zone in Australia (Imber et al., 1991) obtained over 50 percent acceptance of the highest bid in the unedited data.

[6] As demonstrated previously, the log model is more sensitive to small specification and sample changes. Thus, the effects of bid structure would be even greater for the log form.

[7] In order to hold sample sizes constant for the comparisons, we randomly exclude respondents from the original all-bids data set. For example, the small-spills data set only had 324 observations after excluding respondents who received the $1,000 bid. Consequently, we randomly selected 324 observations from the all-bids data set for the comparison.

Table 7-3. Predicted Mean WTP and Standard Errors for Alternative Bid Structures: Dichotomous-Choice Data[a]

	Excluding $1,000 Bid	Comparable with All Bids
Small Spills		
Mean	114	212
Standard error	35	112
Standard error / Mean	0.31	0.53
All Spills		
Mean	125	436
Standard error	34	241
Standard error / Mean	0.27	0.55

[a] Means and standard errors estimated with a linear model.

7.4 Summary and Implications

We have illustrated some effects on dichotomous-choice estimation of choices among several plausible estimation approaches and experimental designs. Our results indicate that the choice of functional form, integration limits, and bid structure among reasonable alternatives can result in widely disparate WTP estimates from the same dichotomous-choice data. All of the alternatives tested are plausible and defensible and have appeared in published studies. Two experienced analysts might well select two different alternatives as the "right" approach for estimating the WTP to prevent oil spills. Unfortunately, the estimates obtained by the two analysts could vary widely. While some variability must be expected in empirical work, using dichotomous-choice formats to estimate nonuse values is not sufficiently reliable for NRDA purposes.

CHAPTER 8

Implications

8.1 Implications for Using CV in NRDAs

Our findings have important implications for using CV to measure natural resource damages, especially for situations in which nonuse damages are likely to be a large share of the total. Our findings are relevant for both developing NRDA regulations and for litigation involving natural resource damages. Specifically, the U.S. Department of the Interior is revising its regulations in response to the Court of Appeals mandate, while the U.S. Department of Commerce is in the early stages of developing regulations for the Oil Pollution Act of 1990. It is likely that CV will continue to be a major issue in litigation for environmental damages, especially for nonuse damages.

Based on our experiments, we find that CV yields estimates that fail to meet several basic criteria for accuracy. Our assessment includes tests of theoretical validity, convergent validity, and some aspects of reliability. The strength of our conclusions lies in the scope, simplicity, and complementarity of our experiments. The scope is broad enough to include several types of natural resource services. The design involves simple, discrete changes in the levels of the resource services to be valued. The design also allows the differences among the questionnaires to be tightly controlled to minimize influences that might confound the comparisons. Additionally, the design involves completely independent tests for two different commodities: preventing deaths of varying numbers of migratory waterfowl and providing different levels of response capability to reduce environmental damages from oil spills. In each experiment, the alternative levels of resource services are completely embedded. For example, the level of response for all oil spills includes that for small oil spills.

Our findings on theoretical validity are based on statistical results from two independent experiments using well-designed questionnaires, large sample sizes, and several different tests. All nonparametric tests and tests using the most plausible parametric models for both commodities reject the hypothesis that CV measures of WTP from our experiments are theoretically valid. The results also are robust across question formats, with tests of both open-ended and dichotomous-choice formats rejecting the hypothesis of theoretical validity.

Additionally, our questionnaires focus on total values (although nonuse values dominate) in an *ex ante* setting—that is, protecting waterfowl and preventing oil-spill damage. Freeman's (1990) criteria indicate that this situation is better-suited for CV than a damage assessment, which is an ex post valuation. The careful experimental design, the extensive pretesting of the questionnaires, the independence of the statistical tests, and the overall consistency of the results across tests substantially reduces the likelihood that the findings are the result of experimental design flaws or are statistical artifacts.

Our tests of convergent validity raise questions about the sensitivity of results for the dichotomous-choice question format. Our comparisons of means and distributions for the question formats reject convergent validity in 5 out of 10 cases. However, four of the five cases where we fail to reject are log models. We have cited evidence that the linear specifications appear to be more appropriate for these data.

Additionally, the unexpected willingness of many respondents to agree to pay large bids (over one-third of respondents saying Yes to a $1,000 bid) may be evidence of starting-point bias that has been mentioned but not demonstrated in previous studies (Randall and Farmer, 1992). People may agree to pay the offered bid because their preference for the commodity is ill-defined and the bid provides a cue about acceptable values. Although the open-ended format could induce respondents to reflect more about their preferences, it also is possible that this format induces a different decision-making heuristic. For example, in pretests respondents often spoke of typical charitable contributions or estimating their fair share of the cost. This behavior may explain why we find many large differences in the mean and median WTP estimates for the open-ended and dichotomous-choice formats, especially when we use the synthetic dichotomous-choice data from the open-ended survey. It also may account for the high levels of variability in the dichotomous-choice estimates.

We also find reliability problems associated with the dichotomous-choice format. The sensitivity of the WTP estimates to the bids that are chosen for the experiment is a potentially serious problem, because the choice of bids requires considerable judgment on the part of analysts. The variability across different models further weakens the performance of the dichotomous-choice models. These findings are especially important because proponents of the dichotomous-choice format argue that it is superior to alternative formats. We can draw no conclusions about the relative superiority of the open-ended and dichotomous-choice formats from our evidence. However, our results confirm those of previous studies indicating that the dichotomous-choice format creates several new difficulties: sensitivity to bid structure, sensitivity to modeling assumptions, and starting-point bias. Additionally, the dichotomous-choice format requires larger sample sizes than open-ended studies, which makes it more costly to implement (see Cameron and Huppert [1991]).

The high variability in the data, which cannot be explained by conventional and even sophisticated modeling efforts, is another important limitation in using CV to measure nonuse damages. The extreme variability—standard deviations that are almost four times the size of the mean WTP in our most extreme case and frequently at least twice the mean—indicates that estimates would not be reliable enough for damage assessments. Of course, some of this variability is attributable to a few influential responses. Nevertheless, there is no generally accepted treatment for dealing with these responses. Outlier analyses, trimmed means, and log transformations all involve some kind of analyst judgment and can have a large effect on the estimates.

Outlier issues and more general reliability and validity concerns actually may be different reflections of the same phenomenon. After spending 9 months working with these data, we are convinced that the data contain information that is not merely random "white noise." Our respondents took the exercise seriously and tried hard to answer the questions. However, it appears that different people solved the valuation problem differently. That is, people employed different strategies to form their answers. Some of these strategies may relate to theoretically correct measures of welfare, while others do not.[1] Currently, there is no technique for analyzing CV data that employ a model of how people answer the valuation question, a deficiency that Smith (1985) noted in his commentary on Hanemann's dichotomous-choice article.

It is useful to compare these results to the Monongahela study (see Desvousges, Smith, and Fisher, 1987), which was a major CV study. That study estimated option price for water quality changes that would affect use and possible use of the Monongahela River. The study showed some sensitivity to question format and substantial variability in the bids. The greater emphasis on nonuse values in our current study accentuates the problems that were observed to a lesser degree for use values from the Monongahela study. Our experiments suggest that people have difficulty determining their value for the types of commodities that are relevant in a damage assessment.

Of course, questions will be raised about our study. The study design includes levels of preventing migratory waterfowl deaths that are small relative to the population of migratory waterfowl in the Central Flyway (only about 2 percent even with 200,000 deaths). It would be possible to argue that respondents' preferences are quite flat over this range and therefore the lack of significant differences is not surprising. Marginal utility is positive, but it is negligible in absolute magnitude. If this argument is true, it implies that CV is not useful for many environmental accidents that affect only a small area or a small percentage of a wildlife population. We think the need to reliably measure the value of these types of marginal changes to nonuse services is one of the most critical issues for using CV for damage assessments. Although the differences in numbers of bird deaths prevented are small, relative to the size of the bird population, these numbers of bird deaths corresponded to three very different oil spills.

Additionally, one might argue that we provide respondents with the wrong information in the oil-spills experiment. We describe the prevention facilities but do not provide detailed descriptions of the service flows provided by the prevention facilities. This also is a legitimate concern. However, in many cases decisions must be

[1] Schkade and Payne (1992) identify several different strategies used by respondents.

made with limited information about likely physical effects. The migratory waterfowl experiment is very specific about the effects. The oil-spill experiment mirrors many *ex ante* nonuse value situations where benefits are more probabilistic and geographically diffuse.

Our experiments also may be criticized for how they were implemented. Undoubtedly, CV practitioners will find ways that the questionnaires could have been done differently. For example, the questionnaires were self-administered and did not involve the exhaustive use of visual aids. We know of no experimental evidence that intensive in-person interviews would eliminate the problems identified in our experiments.

However, we have not been able to identify any features of our mall samples that would confound our experimental tests. Another target of criticism may be that our samples are drawn from malls. Our questionnaires were carefully designed and thoroughly pretested. They are good examples of instruments that experienced CV researchers are likely to produce with reasonable time and budget constraints, consistent with the NRDA regulations.

After careful scrutiny, we have concluded that our results are not merely artifacts of the characteristics of our survey instrument, sample respondents, or the survey administration. Of course, our experiments could be replicated with more representative samples and in-person interviews to ensure that this conclusion is correct. We have no reason to expect our results would change. The Diamond et al. (1992) and McFadden and Leonard (1992) papers report results for telephone surveys of representative samples of respondents. These studies also show that the CV estimates for nonuse values do not correspond to economic theory.

8.2 Conclusion

Our results demonstrate substantial problems in trying to use CV to measure nonuse values in a damage assessment situation. We do not want to imply that there are no potential nonuse damages from oil spills or hazardous substance releases. Wetland areas, for example, may provide habitat services that are reduced as a result of a spill. We are concerned, however, that currently available estimates of these values are not reliable enough to be used in a damage assessment. Some argue that these values are predominantly ethical or moral in nature and cannot be monetized in a meaningful way. However, if nonuse damages are determined to be separately compensable, compensation should be based on actual amounts people would pay for these kinds of services.

Finally, we think we owe our colleagues some reflections on the evolution of our thinking about the problem of measuring nonuse values. We realize some people believe that Exxon's sponsorship of our research hindered our scientific objectivity. In fact, our sponsor imposed very few constraints on our work and never implied that we should retract or reconsider our findings. At the outset, we and some of our colleagues at Exxon believed that it was very difficult to estimate nonuse values accurately. However, we also thought that it could be done with scrupulous attention to detail, sufficient time, and generous funding. After months of listening to conscientious respondents trying to answer difficult questions and intensively analyzing our data, we cannot maintain our initial confidence in using CV for measuring nonuse values. Given the current state of the art in measuring nonuse damages, we think these damage estimates are neither valid nor reliable.

References

Alberini, Anna. 1991a. "Choice of Thresholds for Efficient Binary Discrete Choice Estimations." Unpublished paper, Department of Economics, University of California, San Diego.

Alberini, Anna. 1991b. "Informational Content of Binary Responses." Unpublished paper, Department of Economics, University of California, San Diego.

Allen, Mary J., and Wendy M. Yen. 1979. *Introduction to Measurement Theory*. Monterrey, CA: Brooks/Cole Publishing Co.

Babbie, Earl R. 1979. *The Practice of Social Research*. 2d ed. Belmont, CA: Wadsworth Publishing Company, Inc.

Beggs, S., S. Cardell, and J. Hausman. 1981. "Assessing the Potential Demand for Electric Cars." *Journal of Econometrics* 16:1-19.

Belsley, David A., Edwin Kuh, and Roy E. Welsch. 1980. *Regression Diagnostics*. New York: John Wiley and Sons.

Bennett, J.W. 1984. "Using Direct Questioning to Value the Existence Benefits of Preserved Natural Areas." *Australian Journal of Agricultural Economics* 28:136-152.

Bergstrom, John C., John R. Stoll, and Alan Randall. 1989. "Information Effects in Contingent Markets." *American Journal of Agricultural Economics* 71:685-691.

Bishop, Richard C., and Kevin J. Boyle. 1985. "The Economic Value of Illinois Beach State Nature Preserve." Final report to the Illinois Department of Conservation. Madison, WI: HBRS, Inc.

Bishop, Richard C., and Thomas A. Heberlein. 1979. "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics* 61:926-930.

Bishop, Richard C., and Thomas A. Heberlein. 1986. "Does Contingent Valuation Work?" In *Valuing Environmental Goods: The Contingent Valuation Methods*, edited by R.G. Cummings, D.S. Brookshire, and W.D. Schulze. Totowa, NJ: Rowman and Allanheld.

Bishop, Richard C., and Thomas A. Heberlein. 1990. "The Contingent Valuation Method." In *Economic Valuation of Natural Resources*, edited by Rebecca L. Johnson and Gary V. Johnson. Boulder, CO: Westview Press.

Bishop, Richard C., Thomas Heberlein, and Mary Jo Kealy. 1983. "Contingent Valuation of Environmental Assets: Comparison with a Simulated Market." *Natural Resources Journal* 23:619-634.

Bowker, John M. and John R. Stoll. 1988. "Use of Dichotomous Choice Nonmarket Methods to Value the Whooping Crane Resource." *American Journal of Agricultural Economics* 70:372-381.

Boyle, Kevin J. 1990. "Dichotomous-Choice, Contingent-Valuation Questions: Functional Form Is Important." *Northeastern Journal of Agriculture and Resource Economics* 19:125-131.

Boyle, Kevin J., and Richard C. Bishop. 1987. "Valuing Wildlife in Benefit-Cost Analyses: A Case Study Involving Endangered Species." *Water Resources Research* 23:943-950.

Boyle, Kevin J., and Richard C. Bishop. 1988. "Welfare Measurements Using Contingent Valuation: A Comparison of Techniques." *American Journal of Agricultural Economics* 70:20-28.

Boyle, Kevin J., Richard C. Bishop, and Michael P. Welsh. 1988. "Starting Point Bias in Contingent Valuation Surveys." *Land Economics* 61:188-194.

Brookshire, David S., and Don L. Coursey. 1987. "Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures." *American Economic Review* 77:554-566.

Brookshire, David S., Larry S. Eubanks, and Alan Randall. 1983. "Estimating Option Prices and Existence Values for Wildlife Resources." *Land Economics* 59:1-15.

Brookshire, David S., Alan Randall, and John R. Stoll. 1980. "Valuing Increments and Decrements in Natural Resource Service Flows." *American Journal of Agricultural Economics* 62:478-488.

Brookshire, David S., Mark A. Thayer, William D. Schulze, and Ralph C. d'Arge. 1982. "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches." *American Economic Review* 72:165-177.

Cameron, Trudy Ann. 1988. "A New Paradigm for Valuing Non-Market Goods Using Referendum Data: Maximum Likelihood Estimation of Censored Logistic Regression." *Journal of Environmental Economics and Management* 15:355-379.

Cameron, Trudy Ann, and Michelle D. James. 1987. "Efficient Estimation Methods for 'Closed-Ended' Contingent Valuation Surveys." *Review of Economics and Statistics* 69:269-276.

Cameron, Trudy Ann, and Daniel D. Huppert. 1991. "Referendum Contingent Valuation Estimates: Sensitivity to the Assignment of Offered Values." *Journal of the American Statistical Association* 86:910-918.

Carmines, Edward G., and Richard A. Zeller. 1979. R*eliability and Validity Assessment*. Beverly Hills, CA: Sage Publications.

Carson, Richard T. 1991. "Contingent Valuation and Visibility Reduction in the Grand Canyon." University of California, San Diego and NBER.

Carson, R.T., N.E. Flores, and W.M. Hanemann. 1992. "On the Nature of Compensable Value in a Natural Resource Damage Assessment." Paper presented at the American Economics Association annual meeting in New Orleans, Louisiana.

Carson, Richard T., W. Michael Hanemann, and Raymond Kopp. 1991. "Comments on the United States Department of Interior's Notice of Proposed Rulemaking 43 C.F.R. Part 11 Natural Resource Damage Assessments." June.

Carson, Richard T., W. Michael Hanemann, and Raymond Kopp. 1992. "Valuing Winter Visibility Improvement in the Grand Canyon." Paper presented at the American Economics Association annual meeting in New Orleans, Louisiana.

Cicchetti, Charles J. 1991. "Comments of Dr. Charles J. Cicchetti Before the U.S. Department of the Interior on Notice of Proposed Rulemaking (56 Fed. Reg. 19752) for Natural Resource Damage Assessment Regulations Type B Rule (43 CFR Part 11), Putnam, Hayes, and Bartlett, Inc., Economic and Management Council. July.

"Conservation Rules and Regulations." 1991. *Tex Reg* 16:2523-2524. May 7.

Cook, Philip J., and Daniel A. Graham. 1977. "The Demand for Insurance and Protection: The Case of Irreplaceable Commodities." *Quarterly Journal of Economics* 91:143-156.

Cooper, Joseph, and John Loomis. 1992. "Sensitivity of Willingness-to-Pay Estimates to Bid Design in Dichotomous Choice Contingent Valuation Models." *Land Economics* 68:211-24.

Cragg, John G. 1971. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods." *Econometrica* 39:829-842.

Cummings, Ronald G., and Glenn W. Harrison. 1992. "Identifying and Measuring Nonuse Values for Natural and Environmental Resources: A Critical Review of the State of the Art." Report prepared for the American Petroleum Institute.

Cummings, Ronald G., David S. Brookshire, and William D. Schulze, eds. 1986. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*. Totowa, NJ: Rowman & Allanheld Publishers.

d'Arge, Ralph C., William D. Schulze, and David S. Brookshire. 1980. "Benefit-Cost Valuation of Long-Term Future Effects: The Case of CO_2." Paper presented at the Resources for the Future/National Climate Program Office Workshop, Fort Lauderdale, FL.

Daubert, John T., and Robert A. Young. 1981. "Recreational Demands for Maintaining Instream Flows: A Contingent Valuation Approach." *American Journal of Agricultural Economics* 63:666-76.

Davis, Robert K. 1964. “The Value of Big Game Hunting in a Private Forest.” In *Transactions of the 29th North American Wildlife and Natural Resources Conference*. Washington, DC: Wildlife Management Institute.

Desvousges, William H., and Richard W. Dunford. 1991. “Comments on the Proposed Revisions in the NRDA Rule Pursuant to the 1989 *Ohio v. Interior* Ruling.” Research Triangle Park, NC: Research Triangle Institute.

Desvousges, William H., Richard W. Dunford, and Jean L. Domanico. 1989. “Measuring Natural Resource Damages: An Economic Appraisal.” Report prepared for the American Petroleum Institute, Washington, DC. January.

Desvousges, William H., and V. Kerry Smith. 1988. “Focus Groups and Risk Communication: The ‘Science’ of Listening to Data.” *Risk Analysis* 8:479-484.

Desvousges, William H., V. Kerry Smith, and Ann Fisher. 1987. “Option Price Estimates for Water Quality Improvements: A Contingent Valuation Study for the Monongahela River.” *Journal of Environmental Economics and Management* 14:248-267.

Desvousges, William H., V. Kerry Smith, and Matthew P. McGivney. 1983. “A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvements.” Research Triangle Park, NC: Research Triangle Institute.

Desvousges, William H., V. Kerry Smith, and Hillery H. Rink, III. 1988. “Communicating Radon Risk Effectively: Radon Testing in Maryland.” Prepared for the Office of Policy, Planning, and Evaluation, U.S. Environmental Protection Agency. Research Triangle Park, NC: Research Triangle Institute.

Diamond, Peter A., Jerry A. Hausman, Gregory K. Leonard, and Mike A. Denning. 1992. “Does Contingent Valuation Measure Preferences? Experimental Evidence.” Paper presented at Cambridge Economics, Inc. Symposium titled, “Contingent Valuation: A Critical Assessment,” in Washington, DC.

Dickie, Mark, Ann Fisher, and Shelby Gerking. 1987. “Market Transactions and Hypothetical Demand Data: A Comparative Study.” *Journal of the American Statistical Association* 82:69-75.

Domencich, T., and D. McFadden. 1975. *Urban Travel Demand: A Behavioral Analysis*. Amsterdam: North Holland.

Duffield, John W., and David A. Patterson. 1992. "Field Testing Existence Values: An Instream Flow Trust Fund for Montana Rivers." Paper presented at the American Economics Association annual meeting in New Orleans, Louisiana.

Dunford, Richard W. 1992. "Natural Resource Damages from Oil Spills." In *Innovations in Environmental Policy*, edited by Tom Tietenberg. Hants, England: Edward Elgar Publishing.

Edwards, Steven F. 1988. "Option Prices for Groundwater Protection." *Journal of Environmental Economics and Management* 15:475-487.

Federal Register. 1986. Department of the Interior Natural Resource Damage Assessment Regulations (51 Fed. Reg. 27692).

Federal Register. 1991. Proposed Changes in the Department of the Interior Natural Resource Damage Assessment Regulations (56 Fed. Reg. 19752-19773).

Fischoff, B., and L. Furby. 1988. "Measuring Values: A conceptual Framework for Interpreting Transactions with Special Reference to Contingent Valuation of Visibility." *Journal of Risk Uncertainty* 1:147-184.

Freeman, A. Myrick III. 1990. "Nonuse Values in Natural Resource Damage Assessment." Working paper, Bowdoin College and Resources for the Future, Inc.

General Accounting Office (GAO). 1991. "Coast Guard: Oil Spills Continue Despite Waterfront Facility Inspection Program." Washington, DC. June.

Greenley, Douglas A., Richard G. Walsh, and Robert A. Young. 1981. "Option Value: Empirical Evidence from a Case Study of Recreation and Water Quality." *Quarterly Journal of Economics* 96:657-74.

Hanemann, W. Michael. 1984. "Welfare Evaluations in Contingent Evaluation Experiments with Discrete Responses." *American Journal of Agricultural Economics* 66:332-341.

Hanemann, W.M. 1991. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?" *American Economic Review* 81:635-647.

Haneman, Michael, John Loomis, and Barbara Kanninen. 1991. "Statistical Efficiency of Double-Bounded Dichotomous-Choice Contingent Valuation." *American Journal of Agricultural Economics* 73:1255-1263.

Hausman, J. 1979. "Individual Discount Rates and the Purchase and Utilization of Energy Using Durables." *Bell Journal of Economics* 10:33-54.

Heberlein, Thomas A., and Richard C. Bishop. 1986. "Assessing the Validity of Contingent Valuation: Three Field Experiments." *Science of the Total Environment* 56:99-107.

Heckman, J., and R. Willis. 1976. "A Beta Logistic Model for the Analysis of Sequential Labor Force Participation of Married Women." *Journal of Political Economy* 85:27-58.

Hoehn, John P. and Alan Randall. 1987. "A Satisfactory Benefit-Cost Indicator from CV." *Journal of Environmental Economics and Management* 14:226-247.

Hogarth, Robin M., and Melvin W. Reder. 1986. *Rational Choice: The Contrast Between Economics and Psychology*. Chicago: University of Chicago Press.

Hubert, James J. 1990. "A Briefing of Migratory Bird Losses in the Oil Fields." Unpublished report, U.S. Fish and Wildlife Service, Albuquerque, NM.

Imber, David, Gay Stevenson, and Leanne Wilks. 1991. "A Contingent Valuation Survey of the Kakadu Conservation Zone." Resource Assessment Commission Research Paper No. 3, Vol.1. Canberra, Australia: Australian Government Publishing Service.

Johansson, Per-Olov, Bengt KristrOrn, and Karl Goran Mdler. 1989. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Response Data: Comment." *American Journal of Agricultural Economics* 71:1054-1056.

Johnson, Rebecca L., N. Stewart Bregenzer, and Bo Shelby. 1990. "Contingent Valuation Question Formats: Dichotomous Choice Versus Open-Ended Responses." In *Economic Valuation of Natural Resources: Issues, Theory, and Applications*, edited by Rebecca L. Johnson and Gary V. Johnson. Boulder, CO: Westview Press.

Kahneman, Daniel, and Jack L. Knetsch. 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management* 22:57-70.

Kanninen, Barbara J. 1990. "Optimal Experiment Design in Contingent Valuation Surveys." Working Paper, Department of Agricultural and Resource Economics, University of California, Berkeley.

Kay, David L, Tommy L. Brown, and David J. Allee. 1987. "The Economic Benefits of the Restoration of Atlantic Salmon to New England Rivers." Department of Natural Resources, New York State College of Agriculture and Life Sciences, Cornell University, NY.

Kealy, Mary Jo, John F. Dovidio, and Mark L. Rockel. 1988. "Accuracy in Valuation Is a Matter of Degree." *Land Economics* 64:158-170.

Kelly, Sean. 1990. "More Birds Died in Southwest Oil Pits than in Valdez Spill, Agents Say." *The News and Observer*, Raleigh, NC, April 8, p. 19A.

Knetsch, Jack L. and Robert K. Davis. 1966. "Comparisons of Methods for Recreation Evaluation." in *Water Research*, edited by A. V. Kneese and S.C. Smith. Baltimore, MD: Johns Hopkins University Press.

Kriström, Bengt. 1988. "Discrete and Continuous Valuation Questions: Do They Give Different Answers?" Working paper #90, Dept. of Forest Economics, Swedish University of Agricultural Sciences, Umea, Sweden.

Larson, Douglas M. 1991. "On Measuring Existence Value." Unpublished working paper, Department of Agricultural Economics, University of California, Davis.

Loomis, John B. 1989. "Test-Retest Reliability of the Contingent Valuation Method: A Comparison of General Population and Visitor Responses." *American Journal of Agricultural Economics* 71:76-84.

Loomis, John B. 1990. "Comparative Reliability of the Dichotomous Choice and Open-Ended Contingent Valuation Techniques." *Journal of Environmental Economics and Management* 18:78-85.

Madalla, F.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

Majid, I., J.A. Sinden, and Alan Randall. 1983. "Benefit Evaluation of Increments to Existing Systems of Public Facilities." *Land Economics* 59:377-392.

McCarthy, J. Thomas. 1984. *Trademarks and Unfair Competition*. 2d ed. San Francisco: Bancroft-Whitney Co.

McConnell, Kenneth E. 1990. "Models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation." *Journal of Environmental Economics and Management* 18:19-34.

McFadden, Daniel L., and Gregory K. Leonard. 1992. "Issues in the Contingent Valuation of Environmental Goods: Methodologies for Data Collection and Analysis." Paper presented at Cambridge Economics, Inc. Symposium titled, "Contingent Valuation: A Critical Assessment," in Washington, DC.

Milgrom, Paul R. 1992. "Is Sympathy an Economic Value? Philosophy, Economics, and the Contingent Valuation Method." Paper presented at Cambridge Economics, Inc. Symposium titled, "Contingent Valuation: A Critical Assessment," in Washington, DC.

Milon, J. Walter. 1989. "Contingent Valuation Experiments for Strategic Behavior." *Journal of Environmental Economics and Management* 17:293-308.

Mitchell, Robert C., and Richard T. Carson. 1981. "An Experiment in Determining Willingness to Pay for National Water Quality Improvements." Draft report prepared for U.S. Environmental Protection Agency. Washington: Resources for the Future, Inc.

Mitchell, Robert C., and Richard T. Carson. 1984. "Willingness to Pay for National Freshwater Quality Improvements." Draft report prepared for U.S. Environmental Protection Agency, Washington, DC: Resources for the Future.

Mitchell, Robert Cameron and Richard T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington, DC: Resources for the Future.

Oil Spill Intelligence Report. 1991a. "International Spill Statistics, 1989-1990." 14:March 28.

Oil Spill Intelligence Report. 1991b. "Outline of Preliminary Statistics of Damages to Natural Resources." 14:April 11.

Park, Timothy, John B. Loomis, and Michael Creel. 1991. "Confidence Intervals for Evaluating Benefits Estimates from Dichotomous Choice Contingent Valuation Studies." *Land Economics* 67:64-73.

Patterson, David A., and John W. Duffield. 1991. "Comment on Cameron's Censored Logistic Regression Model for Referendum Data." *Journal of Environmental Economics and Management* 20:275-283.

Phillips, C.V. and R.J. Zeckhauser. 1989. "Contingent Valuation of Damage to Natural Resources: How Accurate? How Appropriate?" *Toxics Law Reporter* 4:520-529.

Randall, Alan, and Michael C. Farmer. 1992. "Starting Price Effects in Contingent Valuation Data Sets." Presented at the W-133 Meetings, Benefits and Costs in Natural Resources Planning, South Lake Tahoe, California.

Randall, Alan, Berry Ives, and Clyde Eastman. 1974. "Bidding Games for Valuation of Aesthetic Environmental Improvements." *Journal of Environmental Economics and Management* 1:132-49.

Reiling, Stephen D., Kevin J. Boyle, Hsiang-tai Cheng, and Marcia L. Phillips. 1989. "Contingent Valuation of a Public Program to Control Black Flies." *Northeastern Journal of Agricultural and Resource Economics* Volume 18.

Reiling, Stephen D., Kevin J. Boyle, Marcia L. Phillips, and Mark W. Anderson. 1990. "Temporal Reliability of Contingent Valuation." *Land Economics* 66:128-134.

Rossi, Peter A., James D. Wright, and Andy B. Anderson. 1983. *Handbook of Survey Research*. New York: Academic Press.

Rowe, Robert D., Ralph C. d'Arge, and David S. Brookshire. 1980. "An Experiment on the Economic Value of Visibility." *Journal of Environmental Economics and Management* 7:1-19.

Rowe, Robert D., William D. Schulze, W. Douglass Shaw, David Schenk, and Lauraine G. Chestnut. 1991. "Contingent Valuation of Natural Resource Damage Due to the Nestucca Oil Spill, Final Report." Prepared for the Department of Wildlife, State of Washington, the British Columbia Ministry of Environment, and Environment Canada. Boulder, CO: RCG/Hagler, Bailly, Inc.

Schkade, David A., and John W. Payne. 1992. "Where Do the Numbers Come From? How People Respond to Contingent Valuation Questions." Paper presented at Cambridge Economics, Inc. Symposium titled, "Contingent Valuation: A Critical Assessment," in Washington, DC.

Schulze, William D., Ralph C. d'Arge, and David S. Brookshire. 1981. "Valuing Environmental Commodities: Some Recent Experiments." *Land Economics* 57:151-73.

Seip, Kalle, and Jon Strand. 1990. "Willingness to pay for environmental goods in Norway: A contingent valuation study with real payment." Paper presented at the American Economics Association annual meeting in Washington, DC.

Sellar, Christine, John R. Stoll, and Jean-Paul Chavas. 1985. "Validation of Empirical Measures of Welfare Change: A Comparison of Nonmarket Techniques." *Land Economics* 61:156-75.

Sellitz, Claire, Lawrence S. Wrightsman, and Stuart W. Cook. 1970. *Research Methods in Social Relations*. 3d ed. New York: Holt, Rinehart and Winston.

Slovic, Paul, Baruch Fischhoff, and Sarah Lichtenstein. 1979. "Rating the Risks." *Environment* 21:(3)14-20.

Smith, V. Kerry. 1984. "A Bound for Option Value." *Land Economics* 60:292-296.

Smith, V. Kerry. 1985. "Some Issues in Discrete Response Contingent Valuation Studies." *Northeastern Journal of Agricultural and Resource Economics* 14:1-4.

Smith, V. Kerry. 1992a. "Arbitrary Values, Good Causes, and Premature Verdicts." *Journal of Environmental Economics and Management* 22:71-89.

Smith, V. Kerry. 1992b. "On Separating Defensible Benefits Transfers from 'Smoke and Mirrors.'" *Water Resources Research* 28:685-694.

Smith, V. Kerry, and William H. Desvousges. 1986a. "Asymmetries in the Valuation for Risk Reductions." *American Economic Review* 76:291-294.

Smith, V. Kerry, and William H. Desvousges. 1986b. *Measuring Water Quality Benefits*. Boston: Kluwer Nijhoff Publishing.

Smith, V. Kerry, and William H. Desvousges. 1987. "An Empirical Analysis of the Economic Value of Risk Changes." *Journal of Political Economy* 95:89-114.

Smith, V. Kerry, William H. Desvousges, and A. Myrick Freeman, III. 1985. "Valuing Changes in Hazardous Waste Risks: A Contingent Valuation Analysis, Volume I." Draft interim report prepared for U.S. Environmental Protection Agency, Benefits Branch, Economic Analysis Division under EPA Cooperative Agreement No. CR-811075.

Smith, V. Kerry, William H. Desvousges, and Ann Fisher. 1986. "A Comparison of Direct and Indirect Methods for Estimating Environmental Benefits." *American Journal of Agricultural Economics* 68:280-290.

Stevens, Thomas H., Jaime Echeverria, Ronald J. Glass, Tim Hager, and Thomas A. More. 1991. "Existence Value of Wildlife." *Land Economics*. 67:390-400.

Sutherland, Ronald J., and Richard G. Walsh. 1985. "Effect of Distance on the Preservation Value of Water Quality." *Land Economics* 61:281-291.

Thayer, Mark A. 1981. "Contingent Valuation Techniques for Assessing Environmental Impacts: Further Evidence." *Journal of Environmental Economics and Management* 8:27-44.

Tversky, Amos, Paul Slovic, and Daniel Kahneman. 1990. "The Causes of Preference Reversal." *American Economic Review* 80:204-217.

United States Department of Transportation (USDOT). 1989. "Polluting Incidents In and Around U.S. Waters." Washington, DC, February.

U.S. Court of Appeals, D.C. Circuit. 1989. State of Ohio v. U.S. Department of the Interior (880 F.2d 432).

U.S. Interagency Committee on Water Resources. 1950. *Proposed Practices for Economic Analysis of River Basin Projects* (the 'Green Book'). Washington, DC: Government Printing Office. Revised, 1958.

Viscusi, W. Kip, and Charles J. O'Connor. 1984. "Adaptive Responses to Chemical Labeling; Are Workers Bayesian Decision Makers?" *American Economic Review* 74:942-956.

Viscusi, W. Kip, Wesley A. Magat, and Joel C. Huber. 1985. "The Effect of Risk Information on Precautionary Behavior." Working Paper No. 85-2.

Walsh, Richard G., John B. Loomis, and Richard A. Gillman. 1984. "Valuing Option, Existence, and Bequest Demands for Wilderness." *Land Economics* 60:14-29.

Water Resources Council. 1983. "Economic and Environmental Principles and Guidelines for Water and Related Land Resources Implementation Studies." Washington, DC: Government Printing Office.

APPENDIX A

Nonuse Values: Some Recent Empirical Studies

Study	Natural Resource	Type of Value Measured	Type of Survey & Year	Sample Size
Bennett (1984)	Nadgee Nature Reserve, Southeastern Australia	Existence value	Personal interview; 1979	Total sample size: 544; 14.2% (77 responses) of the total sample size was invalid; Final sample size: 467.
Bishop and Boyle (1985)	Illinois Beach State Park and Nature Preserve	Use, option, and existence values	Mail survey; 1985	Total sample size: 571; 37.1% (212 responses) of the total sample size was invalid; Final sample size: 359.
Bowker and Stoll (1988)	Whooping crane—migration path between Canada and Texas	Existence value	Mail survey and on-site survey; 1983	Total sample size: 1,031; 28.1% (290 responses) of the total sample size was invalid; Final sample size: 741
Boyle and Bishop (1987)	Two endangered species in Wisconsin: the bald eagle and the striped shiner	Use and existence values	Mail survey; 1984	Total sample size: 810; No mention of invalid responses.

Question Format	Payment Vehicle	Range of Estimates	Comments on Outliers and Protest Bids
Direct question	One-time lump-sum payment	$0 to $750 Mean = $27.08; standard deviation: $68.82; median: $5.21; mode: $0 ($10 if you exclude $0 bids).	Reasons for zero bids given in the report, although protest bids were not eliminated.
Dichotomous choice; Bids ranged from $1 to $77.	Annual membership to a private foundation that would carry out activities for preserving the area	$1 to $77 Weighted average mean: $27.55; median: $16.44.	Series of questions used to identify valid responses.
Dichotomous choice; Bids ranged from $1 to $130.	Annual membership to an independent foundation that would purchase and maintain refuge land	$5 to $149 Mean WTP was $21 - $149 depending upon functional specification. The mean was calculated using a 95% truncation level or using the highest offer amount ($130). Estimated median: $62 - $67.	No mention of outliers or protest bids. Eliminated responses containing omissions.
Dichotomous choice; Bids ranged from $1 and $100.	Annual membership to a private foundation that would be responsible for the activities preserving the species	Existence value for the bald eagle: $4.92 - $28.38/yr. Striped shiner $1.00 - $5.66/yr. Total value: $6.50 - $75.31/yr. The mean was calculated using a 90% truncation level or the highest offer ($100), whichever was larger.	Asked respondents who were unwilling to pay their reasons for a zero bid; chose not remove any of the "no responses" from the data based on the responses to the questions.

Study	Natural Resource	Type of Value Measured	Type of Survey & Year	Sample Size
Brookshire, Eubanks, and Randall (1983)	Grizzly bear and bighorn sheep in Wyoming	Use, option, and existence values	Mail survey; 1983	Total sample size: 751 for the grizzly bear survey; 785 for the bighorn sheep survey. No mention of invalid responses.
Carson (1991)	Visibility at Grand Canyon National Park	Use, option, and existence values	Personal interview; 1990	Total sample size: 202; 9.4% (19 responses) of the total sample size was invalid; Final sample size: 183
Edwards (1988)	Potable supply of groundwater in Cape Cod, MA	Use and bequest values	Mail survey; 1987	Total sample size: 1,000; 41.5% (415 responses) of the total sample size did not provide sufficient information; Final sample size: 585.

Question Format	Payment Vehicle	Range of Estimates	Comments on Outliers and Protest Bids
Direct question	Annual purchase of a grizzly bear (or bighorn sheep) stamp to ensure its future availability. Four different probabilities of availability were used: 90%, 75%, 50%, and 25%.	Option value for the grizzly bear: \$10.00 - \$21.50/yr. bighorn sheep: \$16.65 - 22.90/yr. The mean was estimated using all responses.	No mention of outliers or protest bids.
Direct question	Higher utility bills to pay for installing scrubbers on the power plants that contribute to visibility impairment in the Grand Canyon	\$0 to \$360 for summer and winter total visibility values; Mean WTP = \$27.78, 10% trimmed mean = \$16.15, 5% trimmed mean = \$20.20; standard deviation: \$50.04; median: \$10.00	Protest zeros and "don't knows" were not included in the WTP estimates. 18 outliers were excluded in the 10% trimmed mean and 9 outliers were excluded in the 5% trimmed mean.
Dichotomous choice; Bids ranged from \$10 to \$2,000.	A bond with annual payments.	Option prices ranged from \$0 to \$1,623 depending on probability of future supply. Option value only made up 1-2% of option price. Bequest values increase the option prices by \$248 when the probability of future supply is .25 to \$975 when the probability of future supply is 1.0.	Protest zeros composed 4.3% of total sample size; they were eliminated. No mention of outliers.

Study	Natural Resource	Type of Value Measured	Type of Survey & Year	Sample Size
Greenley, Walsh and Young (1981)	Water quality for recreational use in the South Platte River Basin, CO	Use, option, existence and bequest values	Personal interview; 1976	Total sample size: 202; No mention of invalid responses
Imber, Stevenson, and Wilks (1991)	Kakadu Conservation Zone and National Park, Northern Territory, Australia	Use, option, existence, and bequest values	Personal interview; 1990	Total sample size: 2,561 (including 502 interviews that were conducted in the Northern Territory). 25 responses were invalid because the surveys were incomplete.
Kay, Brown, and Allee (1987)	Atlantic salmon restoration in New England rivers	Use, option, and existence values	Mail survey; 1986	Total sample size: 677

Question Format	Payment Vehicle	Range of Estimates	Comments on Outliers and Protest Bids
Bidding game: the sales tax starting point was one-half cent per dollar of expenditure with incremental changes of one-fourth cent; the water/sewer fee starting point was 50 cents per month with incremental changes of 50 cents per month.	1. Sales tax 2. Water/Sewer fee	Total nonuse value $42/yr. existence value: $25/yr; bequest value: $17/yr. The nonuse mean was estimated using the responses from the 20% who do not use the River Basin. Total nonuse value for present users: $67/yr. existence value: $34/yr.; bequest value: $33/yr.	None of the zero bids were eliminated, although the report does give a breakdown of reasons for zero bids. No mention of outliers.
Dichotomous choice; Bids ranged from $2 to $250	A reduction in take-home pay or other income caused by both the loss of tax revenue from the mine and the need for money to set up and manage the Zone as part of the park	$2 to $250; This study uses median values for WTP estimates. Major impact scenario: $123.80 - $143.20/yr. Minor impact scenario: $52.80 - $80.30/yr.	No mention of outliers or protest bids.
Users: direct question asking WTP beyond the price of a fishing license; Nonusers: dichotomous choice; either 12 cents/kWh or 9 cents/kWh	Increased taxes or electric bills	Mean WTP: Use: $31.93 in addition to the cost of the fishing license; Option value $10.81; Existence value: $27.45	Extensive discussion on handling nonresponse. No mention of outliers.

Study	Natural Resource	Type of Value Measured	Type of Survey & Year	Sample Size
Majid, Sinden, and Randall (1983)	Two additional parks in an existing park system in Armidale, New South Wales, Australia	Use, option, and existence values	Personal interview; 1983	Total sample size: 140; No mention of invalid responses.
Mitchell and Carson (1984)	Water quality for all rivers and lakes in the United States	Use, option, and existence values. Separated nonuse from use by assuming bids of nonusers are representative of nonuse values.	Personal interview; 1981	Total sample size: 813; 30.6% (249 responses) of the total sample size was invalid; Final sample size: 564.
Rowe, Schulze, Shaw, Schenk, and Chestnut (1991)	Natural resource damages caused by the *Nestucca* oil spill off the coast of Washington and British Columbia	Use, option, existence, and bequest values	Mail survey 1990	Total sample size: 2,515; Washington state: 1,291, British Columbia: 1,224; 26% (654 responses) of the total sample size was invalid; Final sample size: 1,861

Question Format	Payment Vehicle	Range of Estimates	Comments on Outliers and Protest Bids
Iterative bidding. No mention of starting point or iterative amount.	An annual contribution to the two additional parks	Park 1: $3.80/yr standard deviation: $5.20/yr. Park 2: $5.30/yr standard deviation: $10.00/yr.	No mention of outliers or protest bids.
Anchored payment cards based on five income categories. Anchor amounts varied according to the tax and spending rates of the respective group	Higher prices and taxes	$93 for boatable water, $70 for fishable water, and $78 for swimmable water. Mean total WTP: $242. Standard error $8/boatable water, $6/fishable water; $9/ swimmable water. Total standard error $19.	Protest zeros were identified through a series of questions. Bids exceeding 5% of a household's income and bids that were less than $5.00 given by people with above average incomes were eliminated.
Payment card with values ranging from $0 to $5,000	Higher prices to pay for programs that prevent one spill over the next 5 years.	Moderate scenario: Washington mean WTP estimates: $65 to $175 (U.S. dollars); British Columbia mean WTP estimates: $45 to $175 (Canadian dollars)	Follow-up questions were asked in the survey to help determine possible protest bids. Protest bids and outliers were eliminated.

Study	Natural Resource	Type of Value Measured	Type of Survey & Year	Sample Size
Schulze and Brookshire (1983)	The Grand Canyon and other national parklands in the Southwest	Existence and use values	Personal interview; 1980	Total sample size: 614; No mention of invalid responses.
Smith and Desvousges (1986b)	Water quality in the Monongahela River Basin, PA	Option price, option, use, and existence values	Personal interview; 1982	Total sample size: 303; 29.7% (90 responses) of the total sample size was invalid: there were 2 no-answer responses, 56 protest bids, and 32 outliers. Final sample size: 213.
Stevens, Echeverria, Glass, Hager, and More (1991)	Four wildlife species recently restored in New England: the bald eagle, Atlantic salmon, wild turkey, and coyote	Existence, bequest, use, and option values (Largely existence values) *Noted that respondents were probably not very familiar with the commodity being valued.	Mail survey; 1989	Total sample size: approximately 750; Protest bids were identified but not eliminated. Final sample size: 508

Question Format	Payment Vehicle	Range of Estimates	Comments on Outliers and Protest Bids
Direct question and bidding game. No mention of starting point or iterative amount	Higher electric power bills for nonusers and higher entrance fees for users in $2 increments over the existing $2 fee	Electric bills: $3.72/mo. to $5.14/mo. for the Grand Canyon only; $6.61/mo. to $9.64/mo. for the entire southwestern parklands region Entrance fees: $3.16/visit to $4.93/visit.	No mention of outliers or protest bids.
Direct question, payment card, and iterative bidding. Two starting points for the bids: $25 bids changed by increments of $5 or $125 bids changed by increments of $10.	Higher taxes and prices for products	$21.00 - $58.00 for users; $14.00 - $53.00 for nonusers; option price for users: $27.00 - $95.00.	Profile of outliers given in the report. Outliers and protest bids were eliminated.
Used six versions of the mail survey. One version used the direct question format. The other five versions used dichotomous choice, the amounts were randomly selected and ranged from $5 to $150.	Donation to a private trust fund set up to protect the wildlife	Mean WTP: bald eagle: $19.28 wild turkey: $11.86 coyote control: $4.20 coyote preservation: $5.35 Atlantic salmon: $7.93 The authors do comment that the estimated WTP increased by 40% when protest responses were eliminated.	Protest bids were identified using follow-up questions, but were not eliminated in the final analysis. No mention of outliers.

Study	Natural Resource	Type of Value Measured	Type of Survey & Year	Sample Size
Sutherland and Walsh (1985)	Water quality in the Flathead Lake and River drainage system, MT	Use, option, existence, and bequest values	Mail survey; 1981	Total sample size: 171; No mention of invalid responses.
Walsh, Loomis, and Gillman (1984)	Wilderness areas in Colorado	Use, option, existence, and bequest values	Mail survey; 1980	Total sample size: 239; 8.7% of the total sample size was invalid: 21 protest zero bids were eliminated, but 8 no-answer responses were included; Final sample size: 218.

Question Format	Payment Vehicle	Range of Estimates	Comments on Outliers and Protest Bids
Direct question	Annual fee to be placed in special fund	Recreation use value: \$7.37; nonuse value: \$56.79 (composed of option value \$10.71, existence value: \$19.88, and bequest value: \$26.37).	No mention of outliers or protest bids.
Direct question	Annual fee to be placed in special fund	Recreation use \$14.00; Nonuse value \$13.92 (composed of option value: \$4.04, existence value: \$4.87, and bequest value: \$5.01) Mean: \$32.00 Mode: \$10.00 - \$19.99 value category. The 95% confidence interval of mean preservation values was equal to ±50%.	No mention of outliers. Protest bids were eliminated.

APPENDIX B

Migratory Waterfowl Questionnaire

Protecting Ducks and Geese: What Is Your Opinion?

INTRODUCTION

In this survey, most of the questions ask about your attitudes and opinions. There are no right or wrong answers and your responses are confidential. Please think carefully about each question and give your best answers. Because we are asking only a few people to answer these questions, your completion of the survey is extremely important.

BACKGROUND ON MIGRATORY WATERFOWL

Ducks and geese are migratory waterfowl, as are several other bird species. Migratory waterfowl nest and mate in the northern parts of North America in the summer months, and fly south to spend the winter months in warmer parts of the continent.

1. How many times in the past 6 months have you **heard** or **read** about issues involving migratory waterfowl? (Please circle the number next to your answer.)

   01 None

   02 One or Two Times

   03 Three or Four Times

   04 Five or Six Times

   05 Seven or More Times

2. Is protecting migratory waterfowl important to you?
(Circle one number and follow that arrow.)

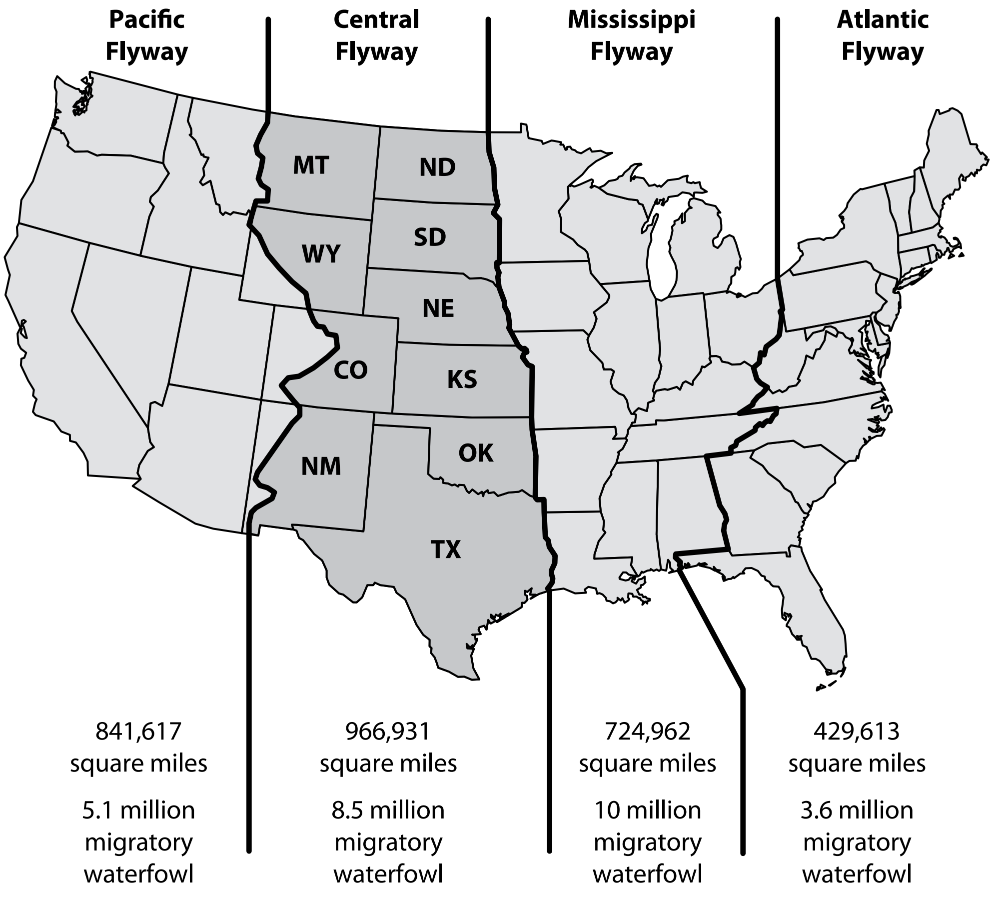
01 YES ↓

02 NO → **GO TO NEXT PAGE**

People have expressed different reasons for protecting migratory waterfowl. Please rate the importance of each of the following reasons for protecting migratory waterfowl. (Circle one number for each statement.)

	Not Important	Somewhat Important	Very Important
My family and I enjoy activities involving migratory waterfowl.	01	02	03
My family and I might enjoy activities involving migratory waterfowl in the future.	01	02	03
Other people enjoy activities involving migratory waterfowl.	01	02	03
Migratory waterfowl are an important part of the ecosystem.	01	02	03
The existence of migratory waterfowl is important.	01	02	03

As shown in the map below, there are four flyways for migratory waterfowl in the continental United States. Waterfowl follow one of these North-to-South pathways when migrating. The **Central Flyway** includes all or part of 10 states in the central United States, from Montana and North Dakota in the north to Texas and New Mexico in the south. It is the largest flyway in land area and has the second highest number of migratory waterfowl (about 8.5 million migratory waterfowl).



3. How would you rate your knowledge of the following threats to migratory waterfowl in the **Central Flyway**? (Circle one number for each threat.)

	Low Knowledge	Medium Knowledge	High Knowledge
Oil Spills	01	02	03
Uncovered Waste-Oil Holding Ponds	01	02	03
Wetlands Destruction	01	02	03
Herbicide/Pesticide Residues	01	02	03

WASTE-OIL HOLDING PONDS

In remote and sparsely populated areas of Texas, Oklahoma, and New Mexico, there are over 250,000 waste-oil holding ponds. These holding ponds contain waste water, oil, and other byproducts from oil and gas drilling operations. They range from 10 - 100 feet in diameter.

Migratory waterfowl are attracted to the waste-oil holding ponds because there are so few wetlands and freshwater ponds in the southern part of the Central Flyway. When they land on the holding ponds, the migratory waterfowl drown because the oil causes them to sink. In 1989, for example, about **20,000** migratory waterfowl died in these holding ponds. This was less than 1% of the 8.5 million migratory waterfowl in the Central Flyway. The affected migratory waterfowl include mallard ducks, pintail ducks, white-fronted geese, snow geese, and greater sandhill cranes.

COVERING WASTE-OIL HOLDING PONDS

To protect migratory waterfowl, the Federal government is considering regulations requiring owners to cover waste-oil holding ponds in the southern part of the Central Flyway. The best type of covering for these holding ponds is a heavy wire netting that prevents waterfowl from landing on the ponds and coming in contact with the oily wastes. The wire netting would be a very small mesh to prevent the waterfowl from getting entangled in the cover.

The proposed regulations would also create a procedure for identifying waste-oil holding ponds and notifying their owners that they must put wire-net covers on their holding ponds. The U.S. Fish and Wildlife Service would monitor the holding ponds to ensure compliance and would cover any abandoned holding ponds. The government would increase taxes on oil and gas producers to pay for the costs of these activities.

4. If the proposed regulations are approved, oil companies would pass on the costs of the wire-net covers to consumers in the form of higher prices. Higher petroleum product prices would, in turn, increase the prices of most other things that you buy.

It is important to know how much protecting these migratory waterfowl is worth to you. Please think about:

- Your current household income
- Your current household expenses
- Other possible uses for your household income

Keeping these factors in mind, what is the **most** that your household would agree to pay **each year** in higher prices for wire-net covers to prevent about 20,000 migratory waterfowl from dying each year in waste-oil holding ponds in the Central Flyway?

$_________ PER YEAR

5. Is your answer to Question 4 **a dollar amount greater than 0?**

(Circle one number and follow that arrow.)

01 YES **02 NO**

Please think about the reason(s) for your dollar amount in Question 4. Read through the entire list below and then put a **1** by the statement that best matches your **most important reason**. (If you have more than one reason, put a 2 by your second most important reason, a 3 by your third most important reason, and so on.)

___ That's how much my household can afford to pay to protect these migratory waterfowl.

___ That's my household's value for protecting these migratory waterfowl.

___ That's my estimate of the cost of wire-net covers per household.

___ That seems like a reasonable amount for my household to pay to protect these migratory waterfowl.

___ That's the amount my household usually contributes to environmental causes.

___ That's more than my household would actually agree to pay, because I want to make sure that these migratory waterfowl are reduced.

___ That's less than my household would actually agree to pay, because I want to make sure that protecting these migratory waterfowl doesn't cost too much.

___ Other (Please describe below.)

____

____

GO TO QUESTION 7

6. Please think about the reason(s) for your response to Question 4. Read through the entire list below and then put a **1** by the statement that best matches **your most important reason**. (If you have more than one reason, put a 2 by your second most important reason, a 3 by your third most important reason, and so on.)

___ My household can't afford to pay any amount to protect these migratory waterfowl.

___ Protecting these migratory waterfowl is not worth any amount to my household.

___ My household would choose to spend its money in other ways.

___ My household should not have to pay to protect these migratory waterfowl.

___ There wasn't enough information for me to answer the question.

___ Higher prices are not a good way to pay for protecting these migratory waterfowl.

___ Wire-net covers would not be effective in protecting these migratory waterfowl.

___ I could not determine a dollar amount for my household for protecting these migratory waterfowl.

___ Other (Please describe below.)

____

____

7. Please indicate the extent to which you agree or disagree with the following statements. (Circle one number for each statement.)

"People have a right to change the environment to meet their needs."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

"People have to make choices between a cleaner environment and a stronger economy."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

"People should preserve the environment at all costs."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

8. Please indicate the extent to which you agree or disagree with the following statements. (Circle one number for each statement.)

"Any death of a migratory waterfowl is a serious problem."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

"Covering waste-oil holding ponds will not significantly affect the migratory waterfowl population in the Central Flyway."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

9. Which of the following activities have you participated in? (Circle all that apply.)

	In the Central Flyway Region*	Elsewhere in the U.S.
Hunting Migratory Waterfowl	01	02
Hunting Wildlife Other than Migratory Waterfowl	01	02
Birdwatching	01	02
Wildlife Viewing (Other than Birdwatching)	01	02
Feeding Birds Around Your Home	01	02

* The Central Flyway includes Colorado, Kansas, Montana, Nebraska, New Mexico, North Dakota, Oklahoma, South Dakota, Texas, and Wyoming.

10. What is your AGE category? (Circle one number.)

01	Under 20
02	20 - 29
03	30 - 39
04	40 - 49
05	50 - 59
06	60 - 69
07	70 or over

11. What is the highest level of EDUCATION you have completed? (Circle one number.)

01	Some high school or less
02	High school diploma
03	Some college
04	College diploma
05	Some graduate school
06	Graduate degree

12. What is your sex? (Circle one number.)

01	Male
02	Female

13. Which of the following categories best describes you? (Circle one number.)

01	Asian-American/Oriental
02	Black/Afro-American/Negro
03	Hispanic-Black/Spanish-speaking Black
04	Hispanic-White/Spanish-speaking White
05	White/Caucasian
06	Native American/American Indian
07	Other (Please specify: ___________________________)

14. Please circle the category below that describes the total amount of INCOME earned by the people in your **household** in 1990.

[Consider all forms of income, including salaries, tips, interest and dividend payments, social security, alimony, and child support.]

01	$15,000 or under
02	$15,001 - $25,000
03	$25,001 - $35,000
04	$35,001 - $50,000
05	$50,001 - $65,000
06	$65,001 - $80,000
07	$80,001 - $100,000
08	Over $100,000

15. How many people live in your household?

____

16. Please indicate if anyone in your household is currently a member of any of the following organizations. (Circle **all** numbers that apply).

01 Nature Conservancy

02 National Geographic Society

03 Audubon Society

04 Sierra Club

05 Ducks Unlimited

06 Cousteau Society

07 National Wildlife Federation

08 Greenpeace

09 World Wildlife Fund

10 Other environmental or conservation organizations (Please specify below.)

____

____

THANK YOU FOR COMPLETING THIS SURVEY

2,000 Bird Version

WASTE-OIL HOLDING PONDS

In remote and sparsely populated areas of Texas, Oklahoma, and New Mexico, there are over 250,000 waste-oil holding ponds. These holding ponds contain waste water, oil, and other byproducts from oil and gas drilling operations. They range from 10 - 100 feet in diameter.

Migratory waterfowl are attracted to the waste-oil holding ponds because there are so few wetlands and freshwater ponds in the southern part of the Central Flyway. When they land on the holding ponds, the migratory waterfowl drown because the oil causes them to sink. In 1989, for example, about **2,000** migratory waterfowl died in these holding ponds. This was much less than 1% of the 8.5 million migratory waterfowl in the Central Flyway. The affected migratory waterfowl include mallard ducks, pintail ducks, white-fronted geese, snow geese, and greater sandhill cranes.

COVERING WASTE-OIL HOLDING PONDS

To protect migratory waterfowl, the Federal government is considering regulations requiring owners to cover waste-oil holding ponds in the southern part of the Central Flyway. The best type of covering for these holding ponds is a heavy wire netting that prevents waterfowl from landing on the ponds and coming in contact with the oily wastes. The wire netting would be a very small mesh to prevent the waterfowl from getting entangled in the cover.

The proposed regulations would also create a procedure for identifying waste-oil holding ponds and notifying their owners that they must put wire-net covers on their holding ponds. The U.S. Fish and Wildlife Service would monitor the holding ponds to ensure compliance and would cover any abandoned holding ponds. The government would increase taxes on oil and gas producers to pay for the costs of these activities.

4. If the proposed regulations are approved, oil companies would pass on the costs of the wire-net covers to consumers in the form of higher prices. Higher petroleum product prices would, in turn, increase the prices of most other things that you buy.

It is important to know how much protecting these migratory waterfowl is worth to you. Please think about:

- Your current household income
- Your current household expenses
- Other possible uses for your household income

Keeping these factors in mind, what is the **most** that your household would agree to pay **each year** in higher prices for wire-net covers to prevent about 2,000 migratory waterfowl from dying each year in waste-oil holding ponds in the Central Flyway?

$________ PER YEAR

200,000 Bird Version

WASTE-OIL HOLDING PONDS

In remote and sparsely populated areas of Texas, Oklahoma, and New Mexico, there are over 250,000 waste-oil holding ponds. These holding ponds contain waste water, oil, and other byproducts from oil and gas drilling operations. They range from 10 - 100 feet in diameter.

Migratory waterfowl are attracted to the waste-oil holding ponds because there are so few wetlands and freshwater ponds in the southern part of the Central Flyway. When they land on the holding ponds, the migratory waterfowl drown because the oil causes them to sink. In 1989, for example, about **200,000** migratory waterfowl died in these holding ponds. This was about 2% of the 8.5 million migratory waterfowl in the Central Flyway. The affected migratory waterfowl include mallard ducks, pintail ducks, white-fronted geese, snow geese, and greater sandhill cranes.

COVERING WASTE-OIL HOLDING PONDS

To protect migratory waterfowl, the Federal government is considering regulations requiring owners to cover waste-oil holding ponds in the southern part of the Central Flyway. The best type of covering for these holding ponds is a heavy wire netting that prevents waterfowl from landing on the ponds and coming in contact with the oily wastes. The wire netting would be a very small mesh to prevent the waterfowl from getting entangled in the cover.

The proposed regulations would also create a procedure for identifying waste-oil holding ponds and notifying their owners that they must put wire-net covers on their holding ponds. The U.S. Fish and Wildlife Service would monitor the holding ponds to ensure compliance and would cover any abandoned holding ponds. The government would increase taxes on oil and gas producers to pay for the costs of these activities.

4. If the proposed regulations are approved, oil companies would pass on the costs of the wire-net covers to consumers in the form of higher prices. Higher petroleum product prices would, in turn, increase the prices of most other things that you buy.

It is important to know how much protecting these migratory waterfowl is worth to you. Please think about:

- Your current household income
- Your current household expenses
- Other possible uses for your household income

Keeping these factors in mind, what is the **most** that your household would agree to pay **each year** in higher prices for wire-net covers to prevent about 200,000 migratory waterfowl from dying each year in waste-oil holding ponds in the Central Flyway?

$________ PER YEAR

APPENDIX C

Open-Ended Oil Spills Questionnaire

Environmental Issues:
What Is Your Opinion?

In this survey, we ask a number of questions about your opinions and attitudes. There are no right or wrong answers and your responses are confidential. Please think carefully about each question and give your best answers. Because we are asking only a few people to answer these questions, your completion of the survey is very important.

1. People have different opinions about the seriousness of various social and economic problems in the United States. Please **rank** the following list of problems in order of their seriousness, with **1** as the **most serious** issue and **6** as the **least serious** issue.

   ___ Drugs and Crime

   ___ Environmental Pollution

   ___ Poor Quality of Education

   ___ Budget Deficit

   ___ Poverty

   ___ High Cost of Health Care

2. There are many sources of environmental pollution in the United States. Unfortunately, the government cannot address all of these sources at the same time. Please indicate whether each of the following pollution sources should be a **low**, **moderate**, or **high** priority for government funding. (Circle only one number for each pollution source.)

	Low Priority	Moderate Priority	High Priority
Solid Waste in Landfills	01	02	03
Air Pollution from Cars and Factories	01	02	03
Radioactive Wastes from Nuclear Power Plants	01	02	03
Water Pollution from Toxic Chemicals	01	02	03
Oil Spills from Tankers	01	02	03
Acid Rain from Power Plants	01	02	03

3. Although there are many environmental issues, the remainder of this survey focuses on oil spills. During the past 6 months, how many times have you **heard** or **read** anything about oil spills in U.S. waters? (Circle the number next to your answer.)

01 None

02 One to Five Times

03 Six to Ten Times

04 Eleven or More Times

Oil is an important source of energy in the United States, representing almost one quarter of our total energy needs. About 75% of all the oil used in the continental United States is transported by tanker ships and barges. Whenever oil is shipped, there is a risk of an oil spill.

4. In 1990, the U.S. Congress passed the Oil Pollution Act to reduce both the chances of any oil spills occurring and the effects of spills on the environment. The Act requires oil companies to:
   - Build new tankers with double hulls and add double hulls to existing tankers
   - Pay higher fines and penalties for spilling oil
   - Upgrade their navigational systems
   - Provide more funds for oil spill research.

   Before receiving this questionnaire, had you **heard** or **read** about the Oil Pollution Act? (Circle one number.)

   01 YES
   02 NO

Most of the Oil Pollution Act focuses on prevention. Experts predict that the Oil Pollution Act will prevent up to 50% of oil spills. Even with the Act, spills will occur. For example, storms can cause spills even when the best management practices are followed. The existing response capabilities are not adequate to control and clean up the oil spills that do occur.

Methods to respond to and clean up oil spills vary according to the size of the spill. About 95% of oil spills in U.S. waters are less than 50,000 gallons. Most of these spills occur in commercial port areas when equipment breaks, malfunctions, or people make mistakes. Some of these spills happen during the loading and unloading of oil from tankers. The remaining spills are the few very large spills that occur offshore or when tankers are entering or leaving port areas. These offshore spills can exceed 1,000,000 gallons.

The extent of the damage from an oil spill depends on its location, the time of year, the weather conditions, and the response actions. The damage from spills of less than 50,000 gallons is usually limited to sea birds and shoreline habitats in the immediate vicinity of the spill. Large offshore spills may cover miles of shoreline, contaminate shellfish beds, and kill thousands of seabirds and some marine mammals. However, wildlife populations quickly recover to their pre-spill levels unless fragile shoreline habitats are damaged.

5. Do you think that it is important to take measures in addition to those included in the Oil Pollution Act to limit the effects of oil spills? (Circle one number.)

   01 YES

   02 NO

To limit the effects of oil spills, cleanup measures must be effective against both the more frequent spills that occur in ports and the few larger spills involving offshore tankers or tankers entering or leaving ports.

Spills Less Than 50,000 Gallons

The effects of oil spills less than 50,000 gallons can be greatly reduced by improving oil-spill response and cleanup capability at ports. The Federal government is considering regulations requiring oil spill response centers at every U.S. port that handles oil. These local response centers will:

- Train local firefighters and other local emergency response personnel in oil spill containment and cleanup techniques.
- Maintain inventories of oil-spill response equipment, such as protective booms, oil skimmers, barges, pumps, and boats.
- Train volunteers to conduct wildlife rescue.

Each local response center will have enough equipment and staff to contain and clean up spills of up to 50,000 gallons quickly and effectively. Experts think that these measures will prevent **90%** of the environmental damage from oil spills of less than **50,000 gallons**.

Spills More Than 50,000 Gallons

Each local response center will respond to any oil spill near its base of operations, but will not be capable of containing or cleaning up most of the oil from large spills of more than **50,000 gallons**.

6. If the proposed regulations are approved, oil companies would have to construct and operate local and regional oil-spill response centers. Oil companies would pass on the costs of the oil-spill response centers to consumers in the form of higher prices. Higher petroleum product prices would, in turn, increase the prices of most other things that you buy.

It is important to know how much reducing the effects of oil spills is worth to you. Please think about:

- Your current household income
- Your current household expenses
- Other possible uses for your household income

Keeping these factors in mind, what is the **most** that your household would agree to pay **each year** in higher prices for local response centers to reduce the effects of oil spills <u>less than</u> **50,000 gallons**?

$________ PER YEAR

7. Is your answer to Question 6 **a dollar amount greater than 0?**

(Circle one number and follow that arrow.)

01 YES **02 NO**

Please think about the most important reason for your dollar amount in Question 6. Read through the entire list below and then circle the one choice that best matches your reason. (Please circle only **one** number.)

01 That's how much my household can afford to pay to reduce the effects of oil spills.

02 That's my household's value for reducing the effects of oil spills.

03 That's my estimate of the cost of oil-spill response centers per household.

04 That seems like a reasonable amount for my household to pay to reduce the effects of oil spills.

05 That's the amount my household usually contributes to environmental causes.

06 That's more than my household would actually agree to pay, because I want to make sure that the effects of oil spills are reduced.

07 That's less than my household would actually agree to pay, because I want to make sure that reducing the effects of oil spills doesn't cost too much.

08 Other (Please describe below.)

____

____

GO TO QUESTION 9

8. Please think about the most important reason for your response to Question 6. Read through the entire list below and then circle the one choice that best matches your reason. (Please circle only **one** number.)

01 My household can't afford to pay any amount to reduce the effects of oil spills.

02 Reducing the effects of oil spills is not worth any amount to my household.

03 My household would choose to spend its money in other ways.

04 My household should not have to pay to reduce the effects of oil spills.

05 There wasn't enough information for me to answer the question.

06 Higher prices are not a good way to pay for reducing the effects of oil spills.

07 Oil-spill response centers would not be effective in reducing the effects of oil spills.

08 I could not determine a dollar amount for my household for reducing the effects of oil spills.

09 Other (Please describe below.)

____

____

9. Based on what you've read or heard about recent oil spills in U.S. waters, please indicate your opinion about the seriousness of the following spills. If you haven't read or heard enough about a particular spill to have an opinion, then circle "Don't Know." (Circle one number for each spill.)

Mega Borg **spill in the Gulf of Mexico off the coast of Texas (July, 1990)**

Not Serious	Somewhat Serious	Very Serious	Extremely Serious	Don't Know
(01)	(02)	(03)	(04)	(05)

American Trader **spill near Huntington Beach, California (February, 1990)**

Not Serious	Somewhat Serious	Very Serious	Extremely Serious	Don't Know
(01)	(02)	(03)	(04)	(05)

Pipeline spill in Arthur Kill between New Jersey and New York (January, 1990)

Not Serious	Somewhat Serious	Very Serious	Extremely Serious	Don't Know
(01)	(02)	(03)	(04)	(05)

Exxon Valdez **spill in Prince William Sound, Alaska (March, 1989)**

Not Serious	Somewhat Serious	Very Serious	Extremely Serious	Don't Know
(01)	(02)	(03)	(04)	(05)

10. Please indicate the extent to which you agree or disagree with the following statements. (Circle one number for each statement.)

"People have a right to change the environment to meet their needs."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

"People have to make choices between a cleaner environment and a stronger economy."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

"People should preserve the environment at all costs."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

11. Please indicate the extent to which you agree or disagree with the following statements. (Circle one number for each statement.)

"Oil spills of any size cause serious environmental damage."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

"Some very large oil spills cause only minimal environmental damage."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

"Most oil spills are the result of oil company negligence."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

12. In a typical year, do you visit the beach or other coastal areas for recreation? (Circle one number.)

01 YES

02 NO

13. What is your AGE category? (Circle one number.)

01	Under 20
02	20 - 29
03	30 - 39
04	40 - 49
05	50 - 59
06	60 - 69
07	70 or over

14. What is the highest level of EDUCATION you have completed? (Circle one number.)

01	Some high school or less
02	High school diploma
03	Some college
04	College diploma
05	Some graduate school
06	Graduate degree

15. What is your sex? (Circle one number.)

01	Male
02	Female

16. Which of the following categories best describes you? (Circle one number.)

01 Asian-American/Oriental
02 Black/Afro-American/Negro
03 Hispanic-Black/Spanish-speaking Black
04 Hispanic-White/Spanish-speaking White
05 White/Caucasian
06 Native American/American Indian
07 Other (Please specify: ___________________________)

17. Please circle the category below that describes the total amount of INCOME earned by the people in your **household** in 1990.

[Consider all forms of income, including salaries, tips, interest and dividend payments, social security, alimony, and child support.]

01 $15,000 or under
02 $15,001 - $25,000
03 $25,001 - $35,000
04 $35,001 - $50,000
05 $50,001 - $65,000
06 $65,001 - $80,000
07 $80,001 - $100,000
08 Over $100,000

18. How many people live in your household?

____

19. Please indicate if anyone in your household is currently a member of any of the following organizations. (Circle **all** numbers that apply).

01 Nature Conservancy
02 National Geographic Society
03 Audubon Society
04 Sierra Club
05 Ducks Unlimited
06 Cousteau Society
07 National Wildlife Federation
08 Greenpeace
09 World Wildlife Fund
10 Other environmental or conservation organizations (Please specify below.)

____

____

THANK YOU FOR COMPLETING THIS SURVEY

Open-Ended, All-Spills Version

To limit the effects of oil spills, cleanup measures must be effective against both the more frequent spills that occur in ports and the few larger spills involving offshore tankers or tankers entering or leaving ports.

Spills Less Than 50,000 Gallons
The effects of oil spills less than 50,000 gallons can be greatly reduced by improving oil-spill response and cleanup capability at ports. The Federal government is considering regulations requiring oil-spill response centers at every U.S. port that handles oil. These local response centers will:

- Train local firefighters and other local emergency response personnel in oil-spill containment and cleanup techniques.
- Maintain inventories of oil-spill response equipment, such as protective booms, oil skimmers, barges, pumps, and boats.
- Train volunteers to conduct wildlife rescue.

Each local response center will have enough equipment and staff to contain and clean up spills of up to 50,000 gallons quickly and effectively. Experts think that these measures will prevent **90%** of the environmental damage from oil spills of less than **50,000 gallons**.

Spills More Than 50,000 Gallons
To combat larger spills, which may exceed 1,000,000 gallons, the regulations will also require regional response centers. These regional centers will be located on the East Coast, Gulf Coast, West Coast, and in Alaska. The regional oil spill response centers will:

- Employ full-time professionals trained in oil spill containment and cleanup
- Maintain inventories of oil spill response equipment designed to deal with large spills
- Provide facilities for large-scale wildlife rehabilitation.

The regional response centers will permit rapid movement of equipment and trained personnel by air to any major spill site in U.S. waters. Experts think that these regional response centers, working with the local response centers, will prevent **75%** of the environmental damage from spills more than **50,000 gallons**.

6. If the proposed regulations are approved, oil companies would have to construct and operate local and regional oil-spill response centers. Oil companies would pass on the costs of the oil-spill response centers to consumers in the form of higher prices. Higher petroleum product prices would, in turn, increase the prices of most other things that you buy.

   It is important to know how much reducing the effects of oil spills is worth to you. Please think about:

   - Your current household income
   - Your current household expenses
   - Other possible uses for your household income

   Keeping these factors in mind, what is the **most** that your household would agree to pay **each year** in higher prices for local and regional response centers to reduce the effects of **all** oil spills?

   $________ PER YEAR

APPENDIX D

Dichotomous-Choice Oil Spills Questionnaire

Environmental Issues: What Is Your Opinion?

In this survey, we ask a number of questions about your opinions and attitudes. There are no right or wrong answers and your responses are confidential. Please think carefully about each question and give your best answers. Because we are asking only a few people to answer these questions, your completion of the survey is very important.

1. People have different opinions about the seriousness of various social and economic problems in the United States. Please **rank** the following list of problems in order of their seriousness, with **1** as the **most serious** issue and **6** as the **least serious** issue.

   ___ Drugs and Crime

   ___ Environmental Pollution

   ___ Poor Quality of Education

   ___ Budget Deficit

   ___ Poverty

   ___ High Cost of Health Care

2. There are many sources of environmental pollution in the United States. Unfortunately, the government cannot address all of these sources at the same time. Please indicate whether each of the following pollution sources should be a **low**, **moderate**, or **high** priority for government funding. (Circle only one number for each pollution source.)

	Low Priority	Moderate Priority	High Priority
Solid Waste in Landfills	01	02	03
Air Pollution from Cars and Factories	01	02	03
Radioactive Wastes from Nuclear Power Plants	01	02	03
Water Pollution from Toxic Chemicals	01	02	03
Oil Spills from Tankers	01	02	03
Acid Rain from Power Plants	01	02	03

3. Although there are many environmental issues, the remainder of this survey focuses on oil spills. During the past 6 months, how many times have you **heard** or **read** anything about oil spills in U.S. waters? (Circle the number next to your answer.)

01 None

02 One to Five Times

03 Six to Ten Times

04 Eleven or More Times

Oil is an important source of energy in the United States, representing almost one quarter of our total energy needs. About 75% of all the oil used in the continental United States is transported by tanker ships and barges. Whenever oil is shipped, there is a risk of an oil spill.

4. In 1990, the U.S. Congress passed the Oil Pollution Act to reduce both the chances of any oil spills occurring and the effects of spills on the environment. The Act requires oil companies to:
    - Build new tankers with double hulls and add double hulls to existing tankers
    - Pay higher fines and penalties for spilling oil
    - Upgrade their navigational systems
    - Provide more funds for oil spill research.

    Before receiving this questionnaire, had you **heard** or **read** about the Oil Pollution Act? (Circle one number.)

    01 YES
    02 NO

Most of the Oil Pollution Act focuses on prevention. Experts predict that the Oil Pollution Act will prevent up to 50% of oil spills. Even with the Act, spills will occur. For example, storms can cause spills even when the best management practices are followed. The existing response capabilities are not adequate to control and clean up the oil spills that do occur.

Methods to respond to and clean up oil spills vary according to the size of the spill. About 95% of oil spills in U.S. waters are less than 50,000 gallons. Most of these spills occur in commercial port areas when equipment breaks, malfunctions, or people make mistakes. Some of these spills happen during the loading and unloading of oil from tankers. The remaining spills are the few very large spills that occur offshore or when tankers are entering or leaving port areas. These offshore spills can exceed 1,000,000 gallons.

The extent of the damage from an oil spill depends on its location, the time of year, the weather conditions, and the response actions. The damage from spills of less than 50,000 gallons is usually limited to sea birds and shoreline habitats in the immediate vicinity of the spill. Large offshore spills may cover miles of shoreline, contaminate shellfish beds, and kill thousands of seabirds and some marine mammals. However, wildlife populations quickly recover to their pre-spill levels unless fragile shoreline habitats are damaged.

5. Do you think that it is important to take measures in addition to those included in the Oil Pollution Act to limit the effects of oil spills? (Circle one number.)

    01 YES

    02 NO

To limit the effects of oil spills, cleanup measures must be effective against both the more frequent spills that occur in ports and the few larger spills involving offshore tankers or tankers entering or leaving ports.

Spills Less Than 50,000 Gallons

The effects of oil spills less than 50,000 gallons can be greatly reduced by improving oil-spill response and cleanup capability at ports. The Federal government is considering regulations requiring oil-spill response centers at every U.S. port that handles oil. These local response centers will:

- Train local firefighters and other local emergency response personnel in oil spill containment and cleanup techniques.
- Maintain inventories of oil spill response equipment, such as protective booms, oil skimmers, barges, pumps, and boats.
- Train volunteers to conduct wildlife rescue.

Each local response center will have enough equipment and staff to contain and clean up spills of up to 50,000 gallons quickly and effectively. Experts think that these measures will prevent **90%** of the environmental damage from oil spills of less than **50,000 gallons**.

Spills More Than 50,000 Gallons

To combat larger spills, which may exceed 1,000,000 gallons, the regulations will also require regional response centers. These regional centers will be located on the East Coast, Gulf Coast, West Coast, and in Alaska. The regional oil-spill response centers will:

- Employ full-time professionals trained in oil-spill containment and cleanup
- Maintain inventories of oil spill response equipment designed to deal with large spills
- Provide facilities for large-scale wildlife rehabilitation.

The regional response centers will permit rapid movement of equipment and trained personnel by air to any major spill site in U.S. waters. Experts think that these regional response centers, working with the local response centers, will prevent **75%** of the environmental damage from spills more than **50,000 gallons**.

6. If the proposed regulations are approved, oil companies would have to construct and operate local and regional oil-spill response centers. Oil companies would pass on the costs of the oil-spill response centers to consumers in the form of higher prices. Higher petroleum product prices would, in turn, increase the prices of most other things that you buy.

   It is important to know how much reducing the effects of oil spills is worth to you. Please think about:

   - Your current household income
   - Your current household expenses
   - Other possible uses for your household income.

   Keeping these factors in mind, would your household agree to pay **$25** more **each year** in higher prices for local and regional response centers to reduce the effects of **all** oil spills? (Circle one number.)

   01 YES
   02 NO

7. Did you answer **YES** to Question 6?

(Circle one number and follow that arrow.)

01 YES **02 NO**

Please think about the most important reason for your YES response to Question 6. Read through the entire list below and then circle the one choice that best matches your reason. (Please circle only **one** number.)

01 My household can afford to pay $25 to reduce the effects of oil spills.

02 My household's value for reducing the effects of oil spills is $25 or more.

03 My estimate of the cost of the oil-spill response centers per household is at least $25.

04 $25 seems like a reasonable amount for my household to pay to reduce the effects of oil spills.

05 My household usually contributes $25 or more to environmental causes.

06 I'm not sure if my household would agree to pay $25 to reduce the effects of oil spills, but I answered YES

07 $25 is more than my household would actually agree to pay, but I said YES because I want to make sure the effects of oil spills are reduced.

07 Other (Please describe below.)

____

____

GO TO QUESTION 9

8. Please think about the most important reason for your NO response to Question 6. Read through the entire list below and then circle the one choice that best matches your reason. (Please circle only **one** number.)

01 My household can't afford to pay $25 to reduce the effects of oil spills.

02 Reducing the effects of oil spills is not worth $25 to my household.

03 My household would choose to spend its money in other ways.

04 My household should not have to pay to reduce the effects of oil spills.

05 Higher prices are not a good way to pay for reducing the effects of oil spills.

06 Oil-spill response centers would not be effective in reducing the effects of oil spills.

07 I'm not sure if my household would agree to pay $25 to reduce the effects of oil spills, but I answered NO.

08 My household would actually agree to pay $25 or more, but I said NO because I want to make sure that reducing the effects of oil spills doesn't cost too much.

09 Other (Please describe below.)

____

____

8a. Would your household agree to pay **any amount** to reduce the effects of **all** oil spills? (Circle one number.)

01 YES

02 NO

9. Based on what you've read or heard about recent oil spills in U.S. waters, please indicate your opinion about the seriousness of the following spills. If you haven't read or heard enough about a particular spill to have an opinion, then circle "Don't Know." (Circle one number for each spill.)

***Mega Borg* spill in the Gulf of Mexico off the coast of Texas (July, 1990)**

Not Serious	Somewhat Serious	Very Serious	Extremely Serious	Don't Know
(01)	(02)	(03)	(04)	(05)

***American Trader* spill near Huntington Beach, California (February, 1990)**

Not Serious	Somewhat Serious	Very Serious	Extremely Serious	Don't Know
(01)	(02)	(03)	(04)	(05)

Pipeline spill in Arthur Kill between New Jersey and New York (January, 1990)

Not Serious	Somewhat Serious	Very Serious	Extremely Serious	Don't Know
(01)	(02)	(03)	(04)	(05)

***Exxon Valdez* spill in Prince William Sound, Alaska (March, 1989)**

Not Serious	Somewhat Serious	Very Serious	Extremely Serious	Don't Know
(01)	(02)	(03)	(04)	(05)

10. Please indicate the extent to which you agree or disagree with the following statements. (Circle one number for each statement.)

"People have a right to change the environment to meet their needs."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

"People have to make choices between a cleaner environment and a stronger economy."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

"People should preserve the environment at all costs."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

11. Please indicate the extent to which you agree or disagree with the following statements. (Circle one number for each statement.)

"Oil spills of any size cause serious environmental damage."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

"Some very large oil spills cause only minimal environmental damage."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

"Most oil spills are the result of oil company negligence."

Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
(01)	(02)	(03)	(04)	(05)

12. In a typical year, do you visit the beach or other coastal areas for recreation? (Circle one number.)

01 YES

02 NO

13. What is your AGE category? (Circle one number.)

01 Under 20
02 20 - 29
03 30 - 39
04 40 - 49
05 50 - 59
06 60 - 69
07 70 or over

14. What is the highest level of EDUCATION you have completed? (Circle one number.)

01 Some high school or less
02 High school diploma
03 Some college
04 College diploma
05 Some graduate school
06 Graduate degree

15. What is your sex? (Circle one number.)

01 Male
02 Female

16. Which of the following categories best describes you? (Circle one number.)

01	Asian-American/Oriental
02	Black/Afro-American/Negro
03	Hispanic-Black/Spanish-speaking Black
04	Hispanic-White/Spanish-speaking White
05	White/Caucasian
06	Native American/American Indian
07	Other (Please specify: ____________________________)

17. Please circle the category below that describes the total amount of INCOME earned by the people in your **household** in 1990.

[Consider all forms of income, including salaries, tips, interest and dividend payments, social security, alimony, and child support.]

01	$15,000 or under
02	$15,001 - $25,000
03	$25,001 - $35,000
04	$35,001 - $50,000
05	$50,001 - $65,000
06	$65,001 - $80,000
07	$80,001 - $100,000
08	Over $100,000

18. How many people live in your household?

____

19. Please indicate if anyone in your household is currently a member of any of the following organizations. (Circle **all** numbers that apply).

01 Nature Conservancy
02 National Geographic Society
03 Audubon Society
04 Sierra Club
05 Ducks Unlimited
06 Cousteau Society
07 National Wildlife Federation
08 Greenpeace
09 World Wildlife Fund
10 Other environmental or conservation organizations (Please specify below.)

____

____

THANK YOU FOR COMPLETING THIS SURVEY

Dichotomous-Choice, Small-Spills Version

To limit the effects of oil spills, cleanup measures must be effective against both the more frequent spills that occur in ports and the few larger spills involving offshore tankers or tankers entering or leaving ports.

Spills Less Than 50,000 Gallons

The effects of oil spills less than 50,000 gallons can be greatly reduced by improving oil-spill response and cleanup capability at ports. The Federal government is considering regulations requiring oil-spill response centers at every U.S. port that handles oil. These local response centers will:

- Train local firefighters and other local emergency response personnel in oil spill containment and cleanup techniques.
- Maintain inventories of oil-spill response equipment, such as protective booms, oil skimmers, barges, pumps, and boats.
- Train volunteers to conduct wildlife rescue.

Each local response center will have enough equipment and staff to contain and clean up spills of up to 50,000 gallons quickly and effectively. Experts think that these measures will prevent **90%** of the environmental damage from oil spills of less than **50,000 gallons**.

Spills More Than 50,000 Gallons

Each local response center will respond to any oil spill near its base of operations, but will not be capable of containing or cleaning up most of the oil from large spills of more than **50,000 gallons**.

6. If the proposed regulations are approved, oil companies would have to construct and operate local and regional oil-spill response centers. Oil companies would pass on the costs of the oil-spill response centers to consumers in the form of higher prices. Higher petroleum product prices would, in turn, increase the prices of most other things that you buy.

It is important to know how much reducing the effects of oil spills is worth to you. Please think about:

- Your current household income
- Your current household expenses
- Other possible uses for your household income.

Keeping these factors in mind, would your household agree to pay **$25** more **each year** in higher prices for local response centers to reduce the effects of oil spills <u>less than</u> **50,000 gallons**? (Circle one number.)

01 YES

02 NO

APPENDIX E

Profile of Outliers

Table E-1. Profile of Outliers: Migratory-Waterfowl Questionnaires

R-Student	D*f*-Fits	WTP ($)	Age[a] (Years)	Income[a] ($)	Education (Year)	Sex	Number of Organizations
2,000 Deaths							
2.61	0.73	2,000	25	90,000	14	Male	5
3.54	0.65	2,000	25	90,000	12	Male	0
3.28	0.65	2,000	25	20,000	10	Male	1
4.12	1.26	2,575	45	90,000	18	Male	5
5.15	1.05	3,000	25	30,000	12	Male	1
6.78	1.32	4,000	25	57,500	12	Male	1
3.31	0.71	2,480	25	42,500	14	Female	2
-2.37	-1.89	5	25	—	16	Male	13
9.31	2.11	5,000	25	30,000	12	Female	1
2.68	0.50	2,000	35	110,000	17	Male	1
20,000 Deaths							
3.38	0.75	1,000	25	—	14	Male	0
2.86	0.65	1,000	25	110,000	18	Male	3
3.01	0.63	1,000	25	57,500	14	Male	0
2.81	0.75	1,000	25	72,500	18	Male	3
3.30	0.66	1,000	25	42,500	10	—	1
2.91	0.46	1,000	35	30,000	14	Male	0
3.02	0.44	1,000	35	7,500	16	Male	0
7.41	1.71	2,000	25	57,500	12	Female	0
2.90	0.61	1,000	25	42,500	12	Male	0
2.75	0.50	1,000	25	—	12	Male	0
3.22	0.52	1,000	25	30,000	12	Male	0
3.30	0.60	1,000	35	42,500	14	Male	0
3.59	0.60	1,000	45	57,500	14	Female	0
2.71	0.87	1,000	25	90,000	14	Male	4
3.13	0.57	1,000	25	20,000	12	Male	0

[a] Midpoint of category.

CONTINUED

Table E-1. Profile of Outliers: Migratory-Waterfowl Questionnaires (continued)

R-Student	D*f*-Fits	WTP ($)	Age[a] (Years)	Income[a] ($)	Education (Year)	Sex	Number of Organizations
200,000 Deaths							
4.92	1.08	2,000	25	30,000	14	Female	0
4.22	1.07	2,000	25	110,000	17	Male	1
3.02	0.59	1,500	25	57,500	12	Male	0
15.94	4.36	5,000	45	57,500	14	Male	1
3.91	1.14	2,000	25	42,500	18	Female	1
2.11	0.49	1,000	55	110,000	16	Male	2

[a] Midpoint of category.

Table E-2. Profile of Outliers: Open-Ended Oil-Spills Questionnaires

R-Student	D*f*-Fits	WTP ($)	Age[a] (Years)	Income[a] ($)	Education (Year)	Sex	Number of Organizations
Small Spills							
1.99	0.45	1,000	25	110,000	14	Male	3
3.63	0.70	2,000	25	42,500	16	Female	1
6.71	1.00	3,000	25	20,000	14	Female	0
5.11	1.30	2,500	25	72,500	16	Male	0
12.25	4.19	5,000	25	110,000	16	Female	1
All Spills							
2.26	0.54	1,000	25	110,000	14	Male	2
-2.24	-1.17	20	25	57,500	14	Male	8
2.35	0.52	1,000	25	57,500	12	Female	0
2.96	0.48	1,000	25	7,500	12	Male	1
1.96	0.62	1,000	35	72,500	17	Female	2
2.08	0.49	1,000	25	42,500	12	Male	0
2.08	0.57	1,000	35	42,500	16	Male	1
9.64	2.83	3,000	25	30,000	14	Female	2
4.92	0.87	1,500	45	9,000	18	Male	0
2.10	0.50	1,000	45	72,500	16	Male	3
2.60	0.85	1,000	45	110,000	18	Female	5
2.04	0.53	1,000	35	57,500	18	Female	0
3.14	0.51	1,000	25	42,500	14	Female	0
4.55	0.77	1,500	25	57,500	12	Male	1

[a] Midpoint of category.

APPENDIX F

General Model Estimates

Chapter 4 outlines a general theoretical model for estimating WTP from open-ended and dichotomous-choice CV data. The utility function in Equations 4.1 and 4.2 allows for a flexible Box-Cox functional form. However, the model estimates reported in the text are limited to the common linear and log special cases widely used in CV studies. These specifications correspond to α equal to zero and one, respectively. Estimates for the general case are not reported in the text because estimates of α do not reliably fall in the theoretically admissible range. Estimates of α that are negative or greater than one invalidate derivation of utility-theoretic mean WTP from the model.

Table F-1 reports α estimates and means for models with the same explanatory variables as the final models shown in the regression tables in Chapter 5.[1] We report estimates for both normal and lognormal distributions.[2] In addition to producing theoretically consistent α estimates, the lognormal distribution is appealing because it constrains WTP to be positive and appears to account for the long tail evident in the dichotomous-choice data.

Lognormal estimates of α for both open-ended and dichotomous-choice data all fall within the theoretically consistent range, although only two estimates are significantly different from zero. Two of the normal estimates are negative for the migratory-waterfowl versions and both normal estimates are greater than one for the dichotomous-choice oil-spills versions. Among the 10 cases where α falls in the theoretically correct range, six of the estimates are not significantly different from zero, suggesting that the linear specification probably is appropriate. The remaining four estimates range from 0.10 to 0.40, again indicating that the linear specification probably is closer to the correct functional form than the log specification.

Unfortunately the lognormal distribution poses a problem in choosing the upper bound of integration for mean WTP. Because of the willingness of a large proportion of dichotomous-choice respondents to accept high bids, the resulting thick tail

[1] See Tables 5-4, 5-5, 5-6, 5-11, 5-12, 5-15, and 5-16.

[2] The lognormal distribution restricts m in Equation 4.4 to be strictly positive. Zero values are accommodated by assuming lognormal is mixed with a degenerate distribution centered at zero. The cumulative distribution function corresponding to Equation 4.8 is

$$F(m;x,\beta,\sigma) = \begin{cases} 1 - \pi & \text{if } m = 0 \quad \text{(F.1a)} \\ 1 - \pi + \pi\Phi\left[\frac{[\log(m) - x\beta]}{\sigma}\right] & \text{if } m > 1 \quad \text{(F.2b)} \end{cases}$$

makes the estimated mean WTP highly sensitive to the choice of integration bound. Table F-1 reports means for five cases ranging from $1,000 (the highest offered bid) to each respondent's income. Similar results have been reported by Bowker and Stoll (1988) for probit and logit models. There appears to be no objective principle for identifying the "correct" upper bound of integration.

While estimates of the general model appear to provide some support for use of the linear functional form, we find no reason to believe that using the general model would affect any of the conclusions reported in the main text.

Table F-1. General Model Estimates

	Normal			Lognormal	
	alpha	Mean WTP		alpha	Mean WTP
Open Ended					
2,000 Birds	0.01	97		0.12	86
20,000 Birds	-0.12*	NA		0.24*	82
200,000 Birds	-0.09	NA		0.15	94
Small Spills	0.10*	162		0.11	172
All Spills	0.19*	110		0.40*	105
Dichotomous Choice					
			Integration Limit		
Small Spills					
	2.69*	NA	Income	0.32	1025
			Income/4		822
			$10,000		611
			$2,000		373
			$1,000		264
All Spills					
	4.08*	NA	Income	0.60	2431
			Income/4		1545
			$10,000		1282
			$2,000		550
			$1,000		349

Notes: (1) Estimated models included the same explanatory variables as the final models reported in Chapter 5.
(2) * indicates the alpha coefficient is statistically significant at the 10% level or better.
(3) NA indicates the mean WTP is not computable because the alpha estimate is either negative or greater than one.

APPENDIX G

Theoretical Validity Tests Using Screened Data

The tables in this Appendix present the results of theoretical validity tests (parametric and nonparametric) using open-ended data from both the migratory waterfowl survey and the oil-spills survey. In contrast to the tests presented in Chapter 5, these tests use data sets that have not had the outliers (as identified by the Belsley, Kuh, and Welsch technique) removed. Protest $0 responses, responses that exceed 25 percent of income, and WTP responses greater than or equal to $10,000 have been removed. We refer to this level of data cleaning as the screened data sets. As shown in the following tables, the results using the screened data sets are consistent with the results found in Chapter 5.

Table G-1. Univariate Results for Three Migratory-Waterfowl Questionnaires: Screened Data Sets

	2,000 Birds	20,000 Birds	200,000 Birds
Mean	162	127	134
Standard deviation	528	257	397
Median	25	25	25
Mode	0	100	100
Range	0 - 5,000	0 - 2,000	0 - 5,000
Shapiro-Wilk statistic[a]	0.33	0.53	0.36
N	298	301	287

[a] This statistic indicates that these distributions are not normal.

Table G-2. Univariate Results for Two Open-Ended Oil-Spills Questionnaires: Screened Data Sets

	Small Spills	All Spills
Mean	190	144
Standard deviation	450	301
Median	50	50
Mode	100	100
Range	0 - 5,000	0 - 3,000
Shapiro-Wilk statistic[a]	0.45	0.51
N	280	296

[a] This statistic indicates that these distributions are not normal.

Table G-3. Nonparametric Tests of Differences in WTP Distributions: Screened Data Sets

Version Comparison	p-Value
Migratory Waterfowl	
2,000 Birds < 20,000 Birds	0.110 ± 0.018
20,000 Birds < 200,000 Birds	0.296 ± 0.026
2,000 Birds < 200,000 Birds	0.209 ± 0.023
Oil Spills (Open-Ended)	
Small Spills < All Spills	0.196 ± 0.023

Table G-4. Migratory Waterfowl WTP Models Based on Tobit Analysis: Screened Data Sets

	Linear Model			Log Model		
	2,000 Birds	20,000 Birds	200,000 Birds	2,000 Birds	20,000 Birds	200,000 Birds
INTERCEPT	20.66 (0.17)	113.62* (1.86)	51.13 (0.53)	0.002 (0.30)	0.004 (0.76)	0.003 (0.52)
READR	-3.10 (0.17)	13.44 (1.61)	28.31** (1.98)	-0.000 (0.40)	0.001** (2.08)	0.001** (2.04)
O_DEADD	156.14** (2.02)	116.70*** (3.30)	35.38 (0.61)	0.006 (1.59)	0.005 (1.52)	0.003 (1.06)
P_HMWA	427.99*** (4.77)	39.04 (0.94)	186.89*** (2.81)	0.010*** (4.19)	-0.000 (0.01)	0.007** (2.30)
AGEYR	-3.25 (1.07)	-4.03*** (2.65)	-2.05 (0.86)	-0.000 (0.69)	-0.000 (1.36)	-0.000 (0.92)
SIGMA	582.03*** (21.21)	270.24*** (21.40)	440.39*** (21.24)	0.018*** (44.74)	0.013*** (56.09)	0.018*** (35.13)
N	276	284	270	276	284	270
Pseudo-R^2	0.09	0.07	0.06	0.06	0.04	0.06
Mean	121	93	99	153	122	149
Standard Error	32	14	20	42	24	32
Z-statistic for Test of Theoretical Validity						
2,000 Birds < 20,000 Birds		0.80			0.66	
20,000 Birds < 200,000 Birds		-0.25			-0.68	
2,000 Birds < 200,000 Birds		0.58			0.09	

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Table G-5. Open-Ended Oil Spills WTP Models Based on Tobit Analysis: Screened Data Sets

	Linear Model		Log Model	
	Small Spills	All Spills	Small Spills	All Spills
INTERCEPT	21.03	-42.16	0.003	-0.000
	(0.40)	(1.17)	(0.81)	(0.07)
READR	21.62**	23.00***	0.000	0.000
	(2.48)	(4.11)	(0.30)	(1.23)
0_EXIST	92.64	52.90	0.003	0.002
	(1.41)	(1.31)	(0.99)	(0.66)
NORGS	20.57	59.60***	0.001	0.001
	(0.82)	(3.49)	(0.64)	(1.21)
SIGMA	484.43***	317.15***	0.020***	0.013***
	(21.76)	(22.02)	(45.41)	(64.21)
N	268	286	268	286
Pseudo-R^2	0.04	0.12	0.01	0.04
Mean	160	120	232	145
Standard Error	24	17	40	25
Z-statistic for Test of Theoretical Validity				
Small Spills < All Spills	1.36		1.84	

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

APPENDIX H

Theoretical Validity Tests Using Trimmed Data (5 Percent and 10 Percent Trims)

The tables in this Appendix present the results of nonparametric theoretical validity tests using open-ended data from both the migratory-waterfowl survey and the oil-spills survey. We have performed these tests on trimmed data sets. To construct these trimmed data sets, first the protest $0 bids are removed. Then some percentage (in this case both 5 and 10 percent) of bids are "trimmed" from either end of the distributions. We provide these results for the sake of completeness. However, we do not feel that there is any theoretical justification for this approach because these distributions are not symmetrically distributed. We provide only the nonparametric test results because arbitrarily removing some portions of respondents and then estimating models seems unjustified.

Table H-1. Univariate Results for Three Migratory-Waterfowl Questionnaires: Trimmed Data Sets

	2,000 Birds	20,000 Birds	200,000 Birds
5% Trim			
Mean	79	92	90
Standard deviation	169	165	159
Median	25	27	25
Mode	100	100	100
Range	0 - 1,000	0 -1,000	0 - 1,000
Shapiro-Wilk statistic[a]	0.44	0.55	0.56
N	273	274	266
10% Trim			
Mean	47	63	62
Standard deviation	51	79	88
Median	25	27	25
Mode	100	100	100
Range	0 - 200	0 - 500	0 - 500
Shapiro-Wilk statistic[a]	0.79	0.71	0.63
N	243	244	236

[a] This statistic indicates that these distributions are not normal.

Table H-2. Univariate Results for Two Open-Ended Oil-Spills Questionnaires: Trimmed Data Sets

	Small Spills	All Spills
5% Trim		
Mean	139	100
Standard deviation	220	160
Median	50	50
Mode	100	100
Range	0 - 1,000	0 - 1,000
Shapiro-Wilk statistic[a]	0.62	0.60
N	257	269
10% Trim		
Mean	104	73
Standard deviation	138	88
Median	50	50
Mode	100	100
Range	0 - 500	0 - 500
Shapiro-Wilk statistic[a]	0.67	0.70
N	229	239

[a] This statistic indicates that these distributions are not normal.

Table H-3. Nonparametric Tests of Differences in WTP Distributions: Trimmed Data Sets

Version Comparison	p-value
Migratory Waterfowl	
5% Trim	
2,000 Birds < 20,000 Birds	0.09 ± 0.02
20,000 Birds < 200,000 Birds	0.39 ± 0.03
2,000 Birds < 200,000 Birds	0.13 ± 0.02
10% Trim	
2,000 Birds < 20,000 Birds	0.06 ± 0.01
20,000 Birds < 200,000 Birds	0.34 ± 0.03
2,000 Birds < 200,000 Birds	0.11 ± 0.02
Oil Spills (Open-Ended)	
5% Trim	
Small Spills < All Spills	0.90 ± 0.02
10% Trim	
Small Spills < All Spills	0.92 ± 0.02

www.rti.org/rtipress **RTI Press publication BK-0001-1009**