Environmental and Resource Economics **25:** 257–286, 2003.

Contingent Valuation and Lost Passive Use: Damages from the Exxon Valdez Oil Spill

RICHARD T. CARSON[1], ROBERT C. MITCHELL[2], MICHAEL HANEMANN[3], RAYMOND J. KOPP[4], STANLEY PRESSER[5] and PAUL A. RUUD[3]

[1]*University of California, San Diego, USA;* [2]*Clark University, USA;* [3]*University of California, Berkeley, USA;* [4]*Resources for the Future, USA;* [5]*University of Maryland, USA*

Accepted 31 March 2003

Abstract. We report on the results of a large-scale contingent valuation (CV) study conducted after the Exxon Valdez oil spill to assess the harm caused by it. Among the issues considered are the design features of the CV survey, its administration to a national sample of U.S. households, estimation of household willingness to pay to prevent another Exxon Valdez type oil spill, and issues related to reliability and validity of the estimates obtained. Events influenced by the study's release are also briefly discussed.

Key words: natural resource damage assessment

JEL classification: Q26

1. Introduction

On the night of 24 March 1989, the Exxon Valdez left the port of Valdez, Alaska and was steaming through the Valdez Narrows on its way to the open waters of Prince William Sound. The tanker left the normal shipping lanes to avoid icebergs from the nearby Columbia Glacier and ran into the submerged rocks of Bligh Reef; its crew failed to realize how far off the shipping lanes the tanker had strayed.[1] Oil compartments ruptured, releasing 11 million gallons of Prudhoe Bay crude oil into the Prince William Sound. It was the largest tanker spill in U.S. waters and to the public it was one of the major environmental disasters in U.S. history.

Prior to the Exxon Valdez oil spill, the estimation of passive use value (Carson, Flores and Mitchell 1999) or as it has often been previously termed, nonuse or existence value, was an area of economic research not well known to many economists working outside the area of benefit cost analysis of projects involving environmental amenities and health risks. However, based on a belief that the State of Alaska and the Federal Government intended to litigate a natural resource damage claim for lost passive use value, the attention paid to the conceptual

underpinnings and estimation techniques for passive use value changed rather abruptly.

Further sparking the rapidly growing interest in passive use values was an important 1989 court opinion, *Ohio v. U.S. Department of the Interior*,[2] which remanded back to the Department of the Interior (DOI) various components of its regulations for conducting natural resource damage assessments under the Clean Water Act and the Comprehensive, Environmental Response, Compensation and Liability Act (CERCLA), commonly known as Superfund.[3] Two particularly important aspects of the court's ruling for passive use value were its findings that: (1) passive use losses were compensable under those Acts and (2) the DOI hierarchy of damage assessment techniques, which placed contingent valuation at the bottom, was unjustified.[4] Interest in passive use values was also heightening at the time of the study by the passage of the Oil Pollution Act of 1990 (OPA) and the regulations that National Oceanic and Atmospheric Administration (NOAA) enacted under it for natural resource damage assessments. The regulations stated: "NOAA believes that the trustee(s) should have the discretion to include passive use values as a component within the natural resource damage assessment determination of compensable values.[5]

This brings us to the current debate over contingent valuation. It is generally recognized that only stated preference methods (Mitchell and Carson 1989; Louviere, Hensher and Swait 2000; Carson, Flores and Meade 2001) are applicable to the estimation of passive use value. Unlike direct use of resources, where for example, one can potentially observe individuals boating and fishing and use these observations to build economic models permitting inference about the value individuals place on such activities,[6] passive use entails no direct involvement with natural resources. As a result, economists are fond of saying passive use leaves no behavioral trace.

Contingent valuation is a survey approach designed to create the missing market for public goods by determining what people would be willing to pay (WTP) for specified changes in the quantity or quality of such goods or, more rarely, what they would be willing to accept (WTA) in compensation for well-specified degradations in the provision of these goods (Hanemann 1999; Bateman et al. 2002).[7] Contingent valuation (CV) circumvents the absence of markets for natural resource services by presenting consumers with a choice situation in which they have the opportunity to buy or sell the services in question. A CV scenario may be modeled after either a private market or a political referendum. The popular name for this form of non-market valuation arose because the elicited values are contingent upon the particular scenario described to survey respondents.

It is fair to say that the debate within the economics community, instigated by the Exxon Valdez spill and the natural resource damage provisions of various laws, includes discussions of both the conceptual underpinnings of passive use and the technique for its measurement. However, it is the measurement technique itself, which has been the target of the sharpest criticism. Much of the recent criticism

of CV is contained in the Exxon-sponsored conference volume, Hausman (1993), and written submissions directed to writers of natural resource damage assessment regulations in DOI and NOAA.[8] To help assess these comments, the NOAA General Counsel, Thomas Campbell, formed a panel of social scientists to explicitly consider the criticisms of contingent valuation and make recommendations to NOAA. The panel was co-chaired by Kenneth Arrow and Robert Solow and was comprised of three additional economists: Edward Leamer of the University of California, Los Angeles, Paul Portney of Resources for the Future and Roy Radner of Bell Laboratories, as well as Howard Schuman, former Director of the Survey Research Center at the University of Michigan. The panel concluded that CV studies convey "useful information" for damage assessment including lost passive use values, provided they follow a number of "stringent guidelines" (Arrow et al. 1993). The recommendations of this panel have influenced the form of both the NOAA and DOI regulations and the wider academic debate.

The results of the CV study conducted for the State of Alaska in preparation for the Exxon Valdez litigation presented here represented the contemporary state-of-the-art, and therefore, stands as a reference point that may be used to assess the criticisms of CV and perhaps the more general debate surrounding passive use. Most of the recommendations made by the NOAA panel to help insure the reliability of CV estimates of lost passive use had already been implemented in the Alaska study including: (1) the use of rigorous probability sampling with a high response rate, (2) in-person interviews, (3) a discrete choice referendum elicitation format, (4) accurate description of the program, (5) conservative design features, (6) checks on understanding and acceptance, (7) debriefing questions following the referendum questions, and (8) careful pretesting. As much of the debate focuses on old CV studies, or small experiments, a reference point portraying CV practice when substantial resources were available to undertake the study should enhance the quality of the debate.[9]

The plan of the paper is as follows. Section 2 discusses the design and development of the survey questionnaire used in the study. Section 3 reviews for the reader the crucial elements of the survey. In section 4 we discuss the execution of the survey including survey sampling, interviewer training, and survey administration. Section 5 presents statistical results, and section 6 contains a postscript on the Exxon Valdez settlement.

2. Survey Design and Development

The Exxon Valdez CV survey instrument was developed over an 18-month period from July 1989 to January 1991. It was designed to be administered, face-to-face, to a national sample. The central part of the survey instrument was the valuation scenario that described the damages caused by the Exxon Valdez oil spill and established a referendum market for eliciting the value respondents place on preventing a future accident that would cause an equivalent amount of

damage in the Prince William Sound area. Other questions preceding and following the scenario asked about the respondent's attitudes, previous awareness of the spill, understanding of the scenario, and personal characteristics. At appropriate places during the in-person interview, display cards, photographs, and maps were shown to the respondent to supplement the information conveyed verbally by the interviewer.

2.1. INITIAL DEVELOPMENT

An extensive program of instrument development research was conducted. The first stage of instrument development involved exploratory research primarily through focus groups. In the second stage, an initial draft questionnaire was produced and revised during a series of one-on-one interviews followed by informal field testing. The third and final stage involved formal field testing and further development work, including a series of four pilot surveys.

The research goal was to develop a valid survey instrument to measure lost passive use values due to the natural resource injuries caused by the Exxon Valdez oil spill. This is a demanding task for the survey designer because the instrument had to meet multiple goals. The first was to measure only a defined set of injuries. This required carefully describing the specific injuries to be valued, the various recovery times for the injured resources, and the available substitutes, to ensure as much as possible that respondents did not value more extensive or less extensive injuries than intended.[10] Open-ended questions at various points in the valuation scenario and diagnostic questions that followed the valuation scenario were used to gauge success in meeting this goal.

The second objective was to ensure consistency with economic theory by eliciting an approximation to the monetized loss in utility suffered by the respondents as a result of the injuries caused by the spill.

The third objective was a basic survey research goal: respondents from all educational levels and varied life experiences should be able to comprehend the language, concepts, and questions used in the survey so that they could make an informed decision. The particular challenge in CV surveys is to convey to respondents what they would get, how it would be provided, and that they would have to pay for it. Given the amount of information it was necessary to convey in the survey, this required an extended period of instrument development research, which is described below. Various diagnostic checks were used in the survey to determine acceptance of scenario features.

Plausibility, the fourth objective, requires that a respondent find the scenario and the payment vehicle believable. Lack of plausibility is a major source of error in CV surveys because it keeps respondents from taking the choice situation seriously. We took various steps to enhance plausibility, including the use of WTP rather than WTA elicitation questions.[11] A referendum format asked respondents to make a

judgment as to whether they would vote for or against a program that, if adopted, would cost their household a specified amount.

The fifth objective was neutrality: the wording and information in the survey instrument should not be perceived by respondents as promoting the interests of any particular party such as the oil companies, government or environmentalists.[12] The instrument's wording was critically peer reviewed at various stages in its development to help assess our success in meeting this objective. A diagnostic question about who the respondents believed was sponsoring the study was also included to see if one party was identified more than another.

The final objective was to be conservative in estimating WTP. When faced with a decision between two wording, design, or analysis options, neither of which was clearly preferred on the basis of theory or solid methodological grounds, we chose the one that would, if it had any effect, lower the aggregate WTP amount. On this basis, for example, pictures of oiled birds were not shown to respondents, a one-time payment was used rather than installment payments, and "don't know" responses were treated as "no" votes instead of dropping them from the sample.

2.2. DESIGN RESEARCH

During the first stage of the instrument development, we conducted six focus groups (Krueger 1988)[13] in different locations around the United States. In the first groups, discussions explored participant knowledge of the Exxon Valdez spill, beliefs about the cause and nature of the harm, and perceptions of the plausibility of possible ways of preventing a future spill. Once particular patterns of understanding and knowledge were established and confirmed, new topics were introduced in subsequent groups.[14]

In the next stage, which took place in the fall of 1989, we developed a draft of the questionnaire and used it to conduct trial interviews. During these in-depth one-on-one interviews, the instrument was repeatedly revised to refine the information it presented and to improve its clarity and flow before experienced interviewers tested it in the field.

During the third stage of instrument development research, which took place from February to November 1990, we conducted four pilot surveys in various parts of the country.[15] The pilots allowed us to test the instrument in a setting close to that of the final survey, obtain quantitative data to assess how the survey instrument was working, and conduct split-sample experiments to investigate key design issues. The location, date, and sample size of the pilot surveys are shown in Table I.[16]

After each pilot, we analyzed the data and revised the questionnaire on the basis of the analysis and extensive interviewer debriefings. The instrument was iteratively revised and improved in this manner until we were confident it met our research objectives.

Table I. Pilot studies for Exxon Valdez study

Pilot I	San Jose, California	February, 1990	N = 105
Pilot II	Dayton & Toledo, Ohio	May, 1990	N = 195
Pilot III	Five rural counties in Georgia	September, 1990	N = 244
Pilot IV	Dayton & Toledo, Ohio	November, 1990	N = 176

2.3. KEY DESIGN ISSUES

Key design issues for this study included the choice of the elicitation method, the nature of the payment vehicle, the years over which payments are collected, and whether the good is valued in a sequence which includes other goods.

With respect to the elicitation method, we determined early in the process that respondents should be asked a binary discrete choice question (Bishop and Heberlein 1979). This type of question, often called a take-it-or-leave-it question, requests the respondent give a yes-or-no response to a specific cost. A single take-it-or-leave-it referendum-like question for a public good is incentive-compatible under fairly general conditions when the government has the ability to compel payment if the policy is implemented; that is, a respondent can do no better than saying "yes" if the policy is actually preferred at the specified cost or by saying "no" if otherwise. The simple binary discrete choice elicitation has been extended to the double-bounded dichotomous choice question (Hanemann, Loomis and Kanninen 1991). Here the respondent is asked to give a yes-or-no response to a second pre-specified higher amount if the response to the initial take-it-or-leave-it question is "yes" and to a pre-specified lower amount if the initial response is "no." Using both responses substantially increases the statistical power of the WTP estimate, i.e., it tends to produce a much tighter confidence interval for the WTP estimate for any fixed sample size; however, it does so at the expense of a downward bias in the estimate because the second response is not, in general, incentive-compatible.[17]

Of the three natural choices for the payment vehicle – higher taxes – higher oil prices, and higher prices over a wider range of goods – only the first two were found to be plausible in our preliminary research. After conducting split-sample comparisons of a tax and an oil price vehicle in Pilots II and III, we decided to use the tax vehicle in the final survey for two reasons. First, the price of gasoline, the major type of oil product through which consumers would pay for the plan if we used the oil prices vehicle, had become quite unstable due to Iraq's invasion of Kuwait. It appeared likely that gasoline prices would increase rapidly in the near future when the final survey would be in the field or, perhaps, decrease if the crisis was resolved peacefully. This instability raised the prospect that if we used the oil prices vehicle, the respondents' WTP amounts might be distorted because of factors unrelated to any economic value they held for preventing future damage

to Prince William Sound. Second, the two split-sample experiments showed that, if anything, the tax vehicle tended to elicit the same (Pilot III) or lower (Pilot II) amounts than those elicited by the oil prices vehicle.

With respect to the number of years over which payments are collected, three major issues were considered. First, longer payment periods mean that budget constraints, particularly for lower income households, are less binding. Second, periodic payments tend to assure respondents that the good will be provided in future years. Third, with multiple year payments some respondents may believe that it is possible for the government to recontract if better opportunities come along. There was no obvious *a priori* basis on which to choose between the lump sum and the annual payment schemes. On the basis of additional focus group work and a telephone survey, we chose the lump-sum payment.[18] Focus group participants were committed to making at least the initial payment and generally to paying for two or three additional years, but any payment schedule longer than that appeared to suffer from the recontracting problem. The lump-sum payment avoids this problem and has the advantage of eliminating the need to determine what rate ought to be applied to discount future payments. It also has the disadvantage of forcing a much tighter budget constraint on respondents, a conservative feature.

Finally, there were two choices related to "embedding."[19] The first was whether to value the good of primary interest by itself or in a sequence of other substitute public goods. Here economic theory provides some important guidance for the valuation of natural resource damages.[20] Due to substitution and income effects, the later in a WTP sequence a normal good is valued, the lower its value. The opposite, however, is true of a WTA sequence; the later in such a sequence a good is valued, the greater its value. These two propositions can be combined with the fact that WTA compensation for a good is greater than or equal to willingness to pay for the same good (Hanemann 1991) to show that valuing a good first (i.e., by itself) in a WTP sequence is the closest approximation to whatever sequence-specific WTA compensation measure is desired (short of being able to measure willingness to accept directly, which is generally difficult to do).

The second "embedding" choice was methodological: the design should ensure that respondents do not answer a different question than the one they are asked, whether by forgetting about their budget constraints or by letting Prince William Sound stand for all oil spills or even all environmental damage. To meet this requirement, the scenario *must* present a plausible choice situation describing the good and its method of provision in adequate detail so that the respondents know what they will and what they will not get. The design choice is whether to value multiple goods in a single survey or to value a single good and carefully differentiate it in the instrument from those other goods with which it might be confused. A survey valuing a single good was used for two reasons. First, it avoids difficulties that are introduced by valuing multiple goods. Second, well-designed single-good CV surveys have been shown to be capable of eliciting values that are sensitive

to the characteristics of the good being valued (Carson and Mitchell 1995; Carson 1997; Carson, Flores and Meade 2001).

In constructing the scenario for this study, we took several steps to minimize the possibility of respondent perceptual error in understanding the good they are being asked to value. First, we paid particular attention in the focus groups and in-depth interviews to how people think about the good we offer them. Second, we used this knowledge, in ways that will be described later, to focus the respondents' attention on what they would and would not get if the program were implemented. Third, each time we used the instrument, both during the development process and in the final interview itself, we asked open- and close-ended questions to assess how well respondents understood what we were attempting to convey in the survey. This enabled us in the analysis to identify the presence of any remaining perceptual problems and, to the extent that they were present, to determine if and how they affected the results.

3. Structure of the Final Questionnaire

CV instruments such as the one used in this study differ from ordinary public opinion surveys in several important respects. One difference is the amount of information that the interviewer conveys to the respondent during the presentation of the scenario. Almost half the length of the 40-minute interview was devoted to informing the respondent about the effects of the spill; a program that could prevent another spill with the same effects, and how the respondent could pay for this program if the respondent thought it was worth the specified cost. A second is the focus of the survey on a single question: whether the respondent would vote for or against the program. The scenario systematically builds up to this question and a series of follow-up questions explores the respondents' reasons for voting the way they did and what they had in mind when they voted. A third is the opportunity respondents were given to change their vote at a later point in the survey in case they wished to do so after further reflection.

To maintain the respondents' interest and enhance their ability to comprehend the information received during their interview, the material was presented in a carefully designed sequence interspersed with visual aids and questions. This section provides an overview of the interview. Boxed or quoted text is from the questionnaire unless otherwise indicated. Survey text in capital letters indicates interviewer instructions not read to the respondents.

3.1. INITIAL QUESTIONS

At the beginning of the interview, respondents did not know that the main subject matter of the survey was the Exxon Valdez oil spill.[21] This allowed us to measure respondent's attitudes about various types of public goods and their prior awareness of the spill before revealing the purpose of the survey.

BOX A

SHOW PHOTO A

This photograph shows Valdez from the air. This is the town (POINT) and across from the town is the terminal where the oil is piped onto tankers (POINT). These are some tankers (POINT).

The tankers go through the narrows here (POINT) into Prince William Sound. The Exxon Valdez tanker went aground on an underwater reef about here (POINT).

This whole area (POINT) is Prince William Sound.

SHOW PHOTO B

The next photo shows a view of part of the Sound.

As you can see, it is ringed with high mountains. In many areas there are glaciers that break up and produce small icebergs. This photo shows the Columbia Glacier which is more than 100 feet high (POINT TO GLACIER WALL). Icebergs from this glacier sometimes float into the shipping lanes.

SHOW PHOTO C

As you can see in the next photo, the area is largely undeveloped.

Most of the land has been set aside as national forest and state parks. People use the area for fishing, boating, camping and other recreation. In the whole area there are only a few small towns. (PAUSE)

After these preliminary questions, the interviewer began to present the elements of the constructed market in which the respondent would later be asked to vote in favor of or against a plan costing the respondent a specific amount of money. This scenario conveyed information about Prince William Sound, the transport of oil by ship from Valdez, the Exxon Valdez spill and its effects, and an escort ship program to prevent damage from another spill that would have the same effect on the environment. At various places during the presentation of the scenario, interviewers showed respondents one of nineteen visual aids: maps, color photographs, and show cards. These materials were designed and pretested to help respondents visualize important aspects of the scenario and to understand the material being read to them.

3.2. PRINCE WILLIAM SOUND DESCRIPTION

After showing respondents a map that located Prince William Sound in the context of Alaska, and Alaska in the context of the United States, the Sound was described in detail with the help of another map. Box A provides a portion of the instrument that illustrates how text and photographs were integrated to covey a sense of the

SHOW PHOTO J

The next photo shows some of the cleanup activity that took place in the summer after the spill. One of the cleanup techniques was to wash as much of the oil as possible off the shore into the water where it was scooped up by special equipment and taken away. It was not possible to remove all the oil from the rocky beaches in this way because some had already soaked into the ground and couldn't be washed out. Scientists believe that natural processes will remove almost all the remaining oil from the beaches within a few years after the spill. (PAUSE)

Sound and its features. Photographs A, B, and C were of various features of the Sound including the Columbia Glacier.

The description then turned to wildlife. During this part of the narrative respondents were shown photographs of living examples of some of the types of wildlife that were killed by the spill. To be conservative, we did not use photographs of actual animals harmed or killed by the spill.

The next section of the scenario described the spill and its impact on the shoreline. After a photograph of a tanker in the sound, the narrative focused on the Exxon Valdez spill. A series of questions were asked at this point to keep the respondent actively involved in the survey. The interviewer then presented a map of the spill area and pointed out where the spill began, how far it traveled, and the time it took for oil to travel that far.

Another map identified the places where the shore was and was not affected in Prince William Sound. Attention was then called to the cleanup effort in the statement, "As you may know, Exxon made a large effort to clean up the oil on the beaches," and in the presentation of Photo J which showed workers washing the oil off a beach. Respondents were given specific information about the duration of the injuries: "Scientists believe that natural processes will remove almost all the remaining oil from the beaches within a few years after the spill."

3.3. DESCRIPTION OF WILDLIFE

The scenario then described the effect of the spill on wildlife. Card 4 displayed information about the twelve bird species most affected by the spill. In addition to the number of dead birds recovered, it gave the total pre-spill population for each of the species to provide a perspective on the available substitutes. For example, with respect to murres, 16,600 were reported dead, and the total population of murres was described as 350,000. Box B presents the narrative that accompanied Card 4. This material communicated a number of important items.

For example, assurance was given that none of these species was threatened with extinction because our focus groups showed that this aspect of the spill injuries was important to respondents. In order to put the bird kill in perspective the text called attention to the fact that large bird kills can occur naturally. Respondents

BOX B

During the period of the spill there were about one and a half million seabirds and sea ducks of various species in the spill area inside and outside Prince William Sound. (POINT)

As you can see from this card, 22,600 dead birds were found. (POINT)

The *actual* number of birds killed by the oil was larger because not all the bodies were recovered. Scientists estimate that the total number of birds killed by the spill was between 75,000 and 150,000.

About *three-fourths* of the dead birds found were *murres*, the black and white bird I showed you earlier. This is shown on the first line of the card. (POINT)

Because an estimated 350,000 murres live in the spill area, this death toll, though high, does *not* threaten the species.

One hundred of the area's approximately 5,000 bald eagles were also found dead from the oil.

The spill did *not* threaten any of the Alaskan bird species, including the eagles, with extinction. (PAUSE)

Bird populations occasionally suffer large losses from disease or other natural causes. Based on *this* experience, scientists expect the populations of all these Alaskan birds to recover within 3 to 5 years after the spill. (PAUSE)

were also told that the numbers of dead birds shown on the cards are limited to those that were recovered and that the actual toll is estimated to be three to six times higher.

Mammal deaths were shown in a table on another card. As with birds, total pre-spill population estimates were provided in addition to kill estimates. Zero kill estimates were listed for three species for which no kills were reported because some pretest respondents had assumed that there were also injuries to these species.

3.4. EXPLANATION OF THE ESCORT SHIP PLAN

The next portion of the scenario introduced the concept of a possible second spill like the Exxon Valdez spill and described how an escort ship program would prevent and/or contain such a spill. It was important that the program be perceived as feasible, effective, and requiring the amount of money later stated as the cost the household would pay if it was approved in the referendum. Respondents were told that if the program were put into effect, two large Coast Guard ships would escort each tanker throughout its journey in Prince William Sound. The escort ships would help prevent an accident and, if an accident occurred, they would keep even a very large spill from spreading beyond the tanker.[22]

BOX C

A-14E. Because everyone would bear *part* of the cost, we are using this survey to ask people how they would vote if they had the chance to vote on the program.

We have found some people would vote *for* the program and others would vote *against* it. Both have good reasons for why they would vote that way.

Those who vote *for* say it is worth money to them to prevent the damage from another large spill in Prince William Sound.

Those who vote *against* mention concerns like the following.

Some mention that it won't protect any other part of the country except the area around Prince William Sound.

Some say that if they pay for this program they would have less money to use for other things that are more important to them.

And some say the money they would have to pay for the program is more than they can afford.

To avoid overburdening the respondents, only information shown in our pretesting to be essential to communicating a plausible choice situation was included in the narrative. For example, mention of the requirement that all tankers should be double-hulled within the next ten years was included because during our pretests we learned that it added credibility. This information also helped to sharply define the ten-year period during which the escort ship program would operate. The narrative further noted that the plan would not provide spill protection outside Prince William Sound.

3.5. VALUATION QUESTIONS

Respondents were informed that the program would be funded by a one-time tax on the oil companies that take oil out of Alaska and that households like theirs would also pay a special one-time federal tax that would go into a Prince William Sound Protection Fund.[23] Immediately before asking the WTP questions, the interviewer presented the material shown in Box C, which was intended to reassure respondents who might not be willing to pay for the program that a no vote was socially acceptable. The reasons presented here for voting against the program came from those given by respondents during the design phase of the research.

The WTP question, A-15, used a discrete-choice referendum elicitation format to ask whether the respondent would vote for the program if it cost a specified amount that would be paid by a one-time federal tax payment. To obtain responses to a range of amounts, four different versions (A through D) of the instrument were

Table II. Program cost by version and question

Version	A-15	A-16	A-17
A	$10	$30	$5
B	$30	$60	$10
C	$60	$120	$30
D	$120	$250	$60

administered to equivalent subsamples. Every respondent was also asked a follow-up amount appropriate to the version they received and their answer to the first WTP question. Those who voted "for" were asked the higher amount for question A-16 and those who voted "against" the lower amount shown for A-17.

The dollar amounts used in this study (see Table II) were based on information about the underlying WTP distribution obtained from the pilot studies.[24] They were chosen to provide reasonable efficiency in estimating key statistics, such as the median, while providing some robustness (Alberini and Carson 1993) with respect to observing a substantially different WTP distribution in the final survey.

The remainder of Section A was devoted to open-ended debriefing questions designed to provide some information about the reasons for respondent answers to the valuation questions. Respondents who said "yes" were asked: "What was it about the program that made you willing to pay for it?" Respondents answering "no" or "not sure" were asked similar questions.

Section B contained a number of questions designed to assess the beliefs respondents held about key elements of the scenario when they answered the WTP questions. Although this type of assessment is difficult to make, it can be very helpful in checking whether respondents understood the scenario and accepted its basic features. Other questions in this section measured attributes that might affect preferences for protecting the Prince William Sound environment from the effects of another oil spill.

In addition to demographic questions, in Section C all respondents who had voted for one or more of the amounts asked about in the WTP questions were asked how strongly they favored the program if it cost this much money. Everyone who answered "not too strongly" or "not at all strongly" was then asked: "All things considered, would you like to change your vote on the program if it cost your household $__ from a vote for the program to a vote against?" The interview concluded with the question that asked for their best guess as to who "employed my company to do this study."[25]

4. Survey Execution

The survey was conducted using a multi-stage area probability sample of residential dwelling units drawn from the 50 United States and the District of Columbia. In the first stage selection, 61 counties or county groups known as primary sampling units (PSU's) were drawn with probabilities proportionate to their population counts.[26] Within these selected PSU's, 334 Census block groups were drawn with probabilities proportionate to their total population counts. The census block groups were stratified by two block characteristics: percent of the population that was black and a weighted average of the value of owner-occupied housing and the rent of renter-occupied housing. In the third stage, approximately 1,600 dwelling units were drawn from the selected blocks. All dwelling units chosen for the sample were then randomly assigned to one of the four different dollar versions of the survey instrument.

A respondent within each dwelling unit was randomly chosen for the interview. After dropping vacant dwelling units and non-English speaking households who were ineligible for the survey,[27] the survey had an overall response rate of 75.2%. This response rate compares favorably with the best academic surveys such as the University of Michigan's American National Election Surveys and the University of Chicago's General Social Survey.

As information about the survey topic was not provided to individuals until the interview was underway, willingness to pay for the Prince William Sound Program *per se* could not have directly affected whether a household responded. It is possible, however, that other characteristics (e.g., household size or, residence in large urban areas) were related to responding/non-responding status. Thus, the composition of the interviewed sample could differ from that of the random sample initially chosen. To help correct this potential problem, sample weights were constructed that incorporated both nonresponse adjustment and poststratification to household totals from the 1990 Decennial Census. The variables used were region, age, race, household size and type (married versus other). Respondents from a western state, the elderly, blacks, and single households tended to be assigned higher weights.

5. Results

5.1. WILLINGNESS TO PAY QUESTIONS

Table III shows the frequencies of each response to question A-15. As expected, the percentage responding with a "for" vote declines as the amount the respondent is asked to pay increases, dropping from 67 percent in favor at $10 to 34 percent at $120.

The A-15 response can be analyzed with a binary discrete choice model, such as a probit, or it can be combined with the A-16 and A-17 responses. Treating the

Table III. A-15 response by version

Version	No	Not sure	Yes
A ($10)	29.92%	2.65%	67.42%
B ($30)	39.33%	8.99%	51.69%
C ($60)	43.53%	5.88%	50.59%
D ($120)	59.14%	6.61%	34.24%

Table IV. Questionnaire version by type of response

Version	Yes-Yes	Yes-No	No-Yes	No-No
A ($10, $30, $5)	45.08%	22.35%	3.03%	29.55%
B (30, 10, 60)	26.04%	26.04%	11.32%	36.60%
C (60, 120, 30)	21.26%	29.13%	9.84%	39.76%
D (120, 250, 60)	13.62%	20.62%	11.67%	54.09%

"not sure" responses as "no" responses results in four response types.[28] These are presented by questionnaire version in Table IV.

The yes-yes and no-no responses are the easiest to interpret because one would expect the yes-yes responses to fall as the dollar amount the respondent is asked to pay goes from $30 in version A (i.e., 45 percent say yes to $30) to $250 in version D (i.e., 14 percent say yes to $250). We would also expect the no-no responses to increase as one moves from version A (i.e., 30 percent say no to $5) to version D (i.e., 54 percent to $60). The no-no responses to version A define the upper bound on the percentage of respondents who may not care about preventing an Exxon Valdez type oil spill. It should be noted, though, that this group of respondents is also likely to include those who do not think that the escort ship plan will work or who believe that the oil companies should pay the entire cost of the plan.

5.2. STATISTICAL MODEL

The type of data gathered using the double-bounded dichotomous choice elicitation method is sometimes referred to as interval-censored survival data (Nelson 1982). Its use in CV work has been explicated at length by Carson and Steinberg (1990), Hanemann, Loomis and Kanninen (1991), Carson, Wilks and Imber (1994), and Haab and McConnell (1997) under the assumption of truthful preference revelation to both questions.[29] Instead of "time", survival is defined with respect to the cost variable. A respondent willing to pay a specific amount "survives" that amount and a respondent who is not willing to pay a specified amount "fails" that amount. A yes-yes response indicates that the respondent's maximum willingness to pay lies

between the A-16 amount and infinity. A yes-no response (i.e., yes to A-15 and no to A-16), indicates that the respondent's maximum WTP amount lies between the amount asked in A-15 and the amount asked in A-16. In survival analysis terms, the failure occurred between the A-15 and A-16 cost amounts. A no-yes response indicates that the respondent's maximum WTP response lies between the amount asked in A-15 and the amount asked in A-17. A no-no response indicates that the respondent's maximum willingness to pay lies between zero and the amount asked in A-17.[30] Thus, a respondent's WTP response can be shown to lie in one of the following intervals depending on the particular response pattern and questionnaire version: A: ($0–$5 [No-No]; $5–$10 [No-Yes]; $10–$30 [Yes-No]; $30–∞ [Yes-Yes]), B: (0–10; 10–30; 30–60; 60–∞), C: (0–30; 30–60; 60–120; 120–∞), D: (0–60; 60–120; 120–250; 250–∞).[31]

The survival analysis framework imposes the key assumption from economic theory that the fraction of the public in favor of the program is weakly monotonically decreasing in its cost. Effectively, the log likelihood function is defined by the difference in WTP density evaluated at two points defined by the two cost amounts the respondent was asked about with the upper end being infinity in the case of a yes-yes response and the lower end being zero in the case of a no-no response. One can maximize this likelihood function assuming a particular parametric distribution, such as the Weibull, or by using Turnbull's (1976) modification of the Kaplan-Meier estimator.[32]

The Turnbull nonparametric approach makes no assumptions about the shape of the underlying WTP distribution. As a result, this technique only estimates the fraction of the density falling into the intervals defined by the different dollar thresholds used in A-15, A-16, and A-17. Table V shows that about 29 percent of the respondents fall into the interval $0 to $5, and that less than 9 percent are willing to pay over $250, and that the median falls into the interval $30–$60.[33] We can also use the estimates of the change in density occurring in each interval to determine a lower-bound estimate for the mean of the WTP distribution. This is done by multiplying the density estimated to be in each interval by the lower endpoint of the interval and then summing over the interval that yields a lower-bound estimate of mean WTP of $53.60.[34] Thus, any empirical distribution that produced the Turnbull interval estimates would result in an empirical estimate of the mean equal to or greater than $53.60.

Maximizing the likelihood function under the assumption of a Weibull distribution yields the estimates in Table VI and result in estimates of $30 for the median and $97 for the mean. The standard errors indicate that the parameters are estimated with reasonable precision and are reflected in the 95 percent confidence intervals for the mean and median. Figure 1 displays the Weibull survival curve. An ideal parametric fit occurs when the parametric survival curve just touches the top of each step of the nonparametric function. The Weibull is a good approximation over most of the dollar range with some indication of divergence in the two tails. This problem can be rectified by fitting a Weibull model that allows for a spike at zero.[35]

Table V. Turnbull estimation results

Lower bound of interval	Upper bound of interval	Probability of being greater than upper bound	Change in density	Asymptotic t-value*
0	5	0.714	0.286	15.46
5	10	0.685	0.029	2.93
10	30	0.535	0.150	10.57
30	60	0.377	0.157	11.04
60	120	0.220	0.157	11.46
120	250	0.088	0.132	9.02
250	∞	0.000	0.088	Normalized
Log-Likelihood – 1325.186			*Against null of no change in density	

Table VI. Weibull estimation results

Parameter	Estimate	Standard error	Asymptotic t-value
Location	58.417	3.914	14.93
Scale	0.558	0.024	23.68
Median $30.30 [$26.18–$35.08]*			Mean $97.18 [$85.82–$108.54]*
Log-Likelihood –1343.014			*95% Confidence Interval

That model significantly improves the fit by placing 20.6% of the respondents at zero and reduces the estimated mean to $79.20 with a 95% confidence interval of [$67.93–$90.47].[36]

5.3. A VALUATION FUNCTION

A valuation function is a statistical way to relate respondents' WTP to their characteristics. They are often estimated to demonstrate the construct validity of the estimate from a CV study. In the simplest sense, the respondent's WTP or an indicator of that WTP is regressed on respondent characteristics such as income and on preferences relevant to the good being valued.

A valuation function is estimated in several steps. First, for observations with missing values in predictor variables, those values must either be imputed or the observations dropped from any estimation using that variable, a generally undesirable option.[37] Next, the variables to include in the valuation function must be determined. Some variables should clearly be included, while for others, the choice

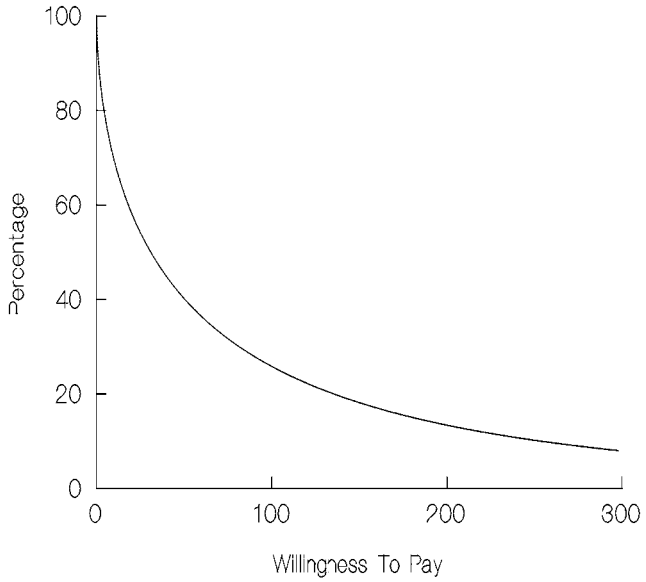

Figure 1. Percent willing to pay as a function of program cost.

is less clear. Finally, the valuation function may be used to make adjustments to WTP estimates for such things as protest responses.

A large number of possible predictors are available for use in the valuation function we wish to estimate. A few, such as income, are obvious choices. Another obvious choice is concern about the environment. Different survey questions that tap this dimension can be used to operationalize this variable in a variety ways. Other good candidates for predictor variables include the likelihood of visiting Alaska and answers to questions that elicit the respondent's perceptions of the characteristics of the oil spill prevention plan. Also, a strong candidate is some indicator of protest responses.

We present our preferred valuation function in Table VII. The first two parameters are the scale and location parameters based on the assumption of a Weibull survival distribution. Note that the scale parameter is a little larger than that estimated in Table VI and the location parameter is quite different because we are parameterizing the original location parameter as a function of the various covariates included in the equation. After the introduction of the covariates, a spike at zero is no longer significant, as many respondents are now predicted to have very small willingness to pay values.

The first four variables, GMORE, MORE, LESS, and NODAM, are dummy variables indicating which respondents believed that the damage likely to occur in the absence of the escort ship plan would be different from that of the Exxon Valdez spill. The coefficients on all four of these variables are significant and follow the expected rank ordering. Those respondents who think that there would be a great

Table VII. Weibull valuation function

Parameter	Estimate	Standard error	Asymptotic *t*-value	Covariate Mean
Location	1.637	1.641	1.00	–
Scale	0.662	0.029	22.91	–
GMORE	0.867	0.284	3.06	0.072
MORE	0.669	0.164	4.07	0.162
LESS	–0.273	0.146	–1.88	0.228
NODAM	–0.794	0.432	–1.84	0.028
MWORK	–0.862	0.131	–6.57	0.265
NWORK	–1.754	0.200	–8.79	0.073
NAME	0.203	0.134	1.51	0.520
COASTAL	0.414	0.143	2.89	0.803
WILD	0.261	0.119	2.19	0.556
STENV	0.473	0.229	2.06	0.098
LIKVIS	0.240	0.138	1.74	0.335
LINC	0.284	0.100	2.85	10.228
WHITE	0.423	0.151	2.80	0.784
PROTEST	–1.226	0.145	–8.45	0.179
Log-Likelihood –1197.728				

deal more damage, GMORE, are willing to pay quite a bit more money than the average respondent. Those who think that there will be somewhat less, but still more damage, MORE, are willing to pay less than the GMORE respondents, but still quite a bit more than the average respondent. Those who think that there would be less damage, LESS, are willing to pay less than the average respondent, and those who think that there would likely be no damage, NODAM, are willing to pay much less. These four variables taken together provide suggestive evidence of respondent sensitivity to the scope of the good valued.

The next two variables, MWORK and NWORK, indicate respondents who think that the plan will prevent less than a great deal of the damage, with MWORK indicating those who think that the plan will prevent some of the damage and NWORK indicating those who think that the plan will not reduce the damage at all. Again, both variables are significant and of the expected negative sign. The NWORK coefficient is about twice the size of the MWORK coefficient in absolute value. The MWORK and NWORK variables provide further evidence undercutting the insensitivity to scope criticism, as they suggest that respondents' valuations and expressed WTP *are* responsive to the characteristics of the good being offered; in this case, the ability of the program to actually prevent the described injuries.

NAME is a dummy variable for those respondents who spontaneously named the Exxon Valdez spill in question A-2 as one of the major environmental accidents caused by humans. As expected, this variable, which measures salience, has a positive influence on a respondent's willingness to pay. COASTAL, which is a dummy variable indicating which respondents said that protecting coastal areas from oil spills was "extremely important" or "very important" in A-3f, has a large and highly significant positive influence on a respondent's willingness to pay. Likewise, WILD, which is a dummy variable for those who felt that the government should set aside a "very large amount" or "large amount" of new land as wilderness in A-4, has a positive effect on a respondent's willingness to pay. STENV, identification of oneself as a strong environmentalist, and LIKVIS, a dummy variable for indicating that the household was "very likely" or "somewhat likely" to visit Alaska in the future, also suggest higher willingness to pay.

Respondents with higher incomes, LINC, are strongly associated with having a higher willingness to pay to prevent another Exxon Valdez type oil spill as is being WHITE. LINC is even more strongly associated with willingness to pay using the subset of respondents for whom income is not imputed. Only 3 of 1043 respondents said "yes" to an amount more than 2% of their income and only 17 said "yes" to an amount more than 1% of their income. Respondents who spontaneously protested (PROTEST) in A-14D or A-15A that Exxon should pay all of the escort ship plan costs (before being asked why they were not willing to pay in A-18), were on average willing to pay much less than those respondents with the same characteristics who did not protest (that Exxon should pay) by this point in the questionnaire.

Depending on a respondent's characteristics, the conditional median willingness to pay predicted by the valuation function varies widely; the lowest predicted value for a respondent in our sample is less than $1 and the highest is $421.

5.4. ADJUSTMENTS AND SENSITIVITY ANALYSIS

The valuation function estimated above allows us to examine the effect that various adjustments would have on our median WTP estimate. The first type of adjustment corrects for respondent assumptions inconsistent with three important features of the scenario. Our information about these inconsistencies comes from respondent answers to questions in Section B of the survey concerning what they had in mind when they answered the WTP questions. Ideally, respondents would have based their WTP amounts on preventing damages of the same magnitude as those caused by the Exxon Valdez spill. For those respondents who did not, there are four dummy variables in our valuation function. One of these has a value of one to represent the particular deviation from this desired perception of the same damage: GMORE, MORE, LESS, and NODAM. Setting the value of these dummy variables to zero effectively forces the perceptions to the same damages. This adjustment reduces the estimate of the median household willingness to pay from $30 to $27.

Another possible adjustment is for the perceived effectiveness of the escort ship plan. Ideally, all respondents would have perceived the plan as being completely effective. One of two dummy variables in the valuation function has a value of one if a respondent indicated that the plan was not completely effective: MWORK and NWORK. Setting both of these dummy variables to zero forces the perception that the plan was completely effective. This adjustment changes the estimate of the median willingness to pay from $30 to $42.

A third adjustment is that for protest responses. The problem here is how to exactly define a protest response. The most conservative definition is the one used in the variable PROTEST in the valuation function. This indicator variable takes the value of one if the respondent volunteered that Exxon or the oil companies should pay before the respondent was asked why he was against the plan (A-18) and takes the value zero otherwise. Setting PROTEST to zero forces out that consideration and changes the estimate of the median from $30 to $37. Making all three adjustments simultaneously yields a point estimate of $48 for the median household willingness to pay to prevent an Exxon Valdez type oil spill.

We have also examined the sensitivity of the estimates to four other factors. The first of these is using only the A-15 response (rather than A-15, A-16 and A-17) since the second response may introduce some bias. Here the Turnbull estimator still places the median in the $30–$60 interval and results in a lower-bound estimate of the mean which is less than $2 lower.[38] Parametric approaches tend to result in somewhat higher estimates. The second is to drop respondents from the sample who may not have clearly understood the CV scenario posed to them. Here a number of more or less inclusive criteria can be employed based upon the interviewer evaluations and responses to particular debriefing questions. In all instances, dropping these respondents raises WTP estimates for the remaining sample. The third is to look at the sponsor question. A plurality (42%) of the respondents believed that Exxon or the oil companies sponsored the survey with the government (23%) being next. Believing that Exxon or the oil companies had sponsored the survey was not a statistically significant predictor of respondent willingness to pay.

The fourth type of sensitivity analysis is a consideration of how stable the estimates of the WTP distribution are over time. This can be done by looking at the surveys completed in Dayton-Toledo, Ohio, two pilot studies and a tracking survey conducted simultaneously with the final survey. This comparison shows that the estimates of the WTP distribution were statistically indistinguishable at three different points in time over the course of a year. The University of Chicago's National Opinion Research Center administered the final questionnaire nationally two years later and, from that data, we obtained almost identical estimates to those reported here (Carson et al. 1997).

5.5. AGGREGATE LOST PASSIVE USE VALUE

The original study reported an estimate of $2.8 billion (1990) dollars as the lower bound on the estimated aggregate lost passive use values. This estimate was obtained by multiplying the number of English-speaking households, the population sampled, by the estimate of median WTP. This estimate was very conservative in two main ways. First, from a theoretical perspective, mean WTA (which is greater than mean WTP) is the most appropriate measure of the services lost or disrupted by the Exxon Valdez oil spill.[39] Second, median WTP is less than the mean WTP under the weak assumption that the WTP distribution is positively skewed. The advantage of the median is that it tends to be quite robustly estimated in survival models and is relatively insensitive to distributional assumptions. Since the report, substantial progress has been made on estimating non-parametric and more flexible parametric models of the WTP distribution. If one were to employ the most conservative estimate of mean WTP consistent with the non-parametric Turnbull density parameters, the estimate of aggregate lost passive use is 4.87 billion dollars.[40] Using the mean WTP estimate from the parametric three-parameter Weibull distribution yields an estimate of 7.19 billion dollars.

These amounts reflect the public's willingness to pay to prevent another Exxon Valdez type oil spill given the scenario posed. Simultaneously adjusting the WTP estimates for protest responses, perceptions of damages larger or smaller than the Exxon Valdez spill, and for perceptions that the proposed plan would not be completely effective, results in higher estimates.

6. A Postscript

The State of Alaska and the U.S. Government settled their lawsuits against Exxon for 1 billion dollars in natural resource damages and restitution for injuries.[41] In addition, Exxon spent over 2 billion dollars on oil spill response and restoration. This compares to the 2.8 billion dollars to prevent an Exxon Valdez type oil spill put forth in the original study report. In thinking about the settlement, it may be useful to keep in mind that guidelines on natural resource damage assessment require that any money collected by the government be spent on restoration and/or the acquisition of like resources where restoration is not feasible. It is clearly possible to argue about which Exxon expenditures represented response (not to be counted toward compensable damages) and which represented restoration (counted toward compensable damages). It is also possible to be critical of the restoration effort. Much, however, has been learned since the Exxon Valdez oil spill about the effects of oil spills, how to prevent them, how to respond to them.[42]

Indeed, instances where a spill is averted receive little attention. After the Exxon Valdez oil spill, the U.S. Coast Guard put into effect an oil spill prevention and response program that strongly resembled the program described to respondents in this study. Their regulatory impact assessment for this plan was based on preventing damages of the magnitude indicated by Exxon's settlement with

the government. The costs of this program have subsequently been passed on to consumers throughout the United States in the form of higher oil prices. After the plan was put into effect, a tanker had problems with its steering system after leaving Valdez and was about 100 feet from hitting the rocks when its escort ship succeeded in pushing it away (Fararo 1992). This use of the study results for a benefit-cost assessment of a program to protect *ex ante* the natural resources of Prince William Sound complete the circle between the usual policy analysis and natural resource damage assessment.[43]

The debate over CV measures of passive use and their role in the assessment of natural resource damages and public decision-making has become a major topic of debate for the economics community (Carson, Flores and Meade 2001). The Exxon Valdez represented the quintessential case in which, to ignore passive use values, was to effectively say that resources that the public had chosen to set aside and not develop could be harmed at little or no cost to the responsible party.

It is possible to belief that lost passive use values should be compensated but not believe in using direct monetary valuation via CV. Requiring restoration of an injured resource as many critics of using monetary valuation had argued should be the remedy has been shown to be a vacuous concept when large numbers of animals are killed and ecosystems disrupted for years. While it is clearly possible to compensate the public by providing additional natural resources to compensate for the lost service flows until the resource recovers, determining the level of compensatory resources that would make the public whole effectively requires knowledge of how much monetary value the public placed on the resource (Flores 2002).

At the time of the Exxon Valdez oil spill it was not clear whether Admiralty law which limits damages to the value of the ship and its cargo would take precedent in determining liability over federal/state pollution statues. The passage of the U.S. Oil Pollution Control Act of 1990 removed that ambiguity and came down clearly on the side of including passive use in assessing damages. That policy decision has not been decisively made elsewhere in the world. As such, perceived liability for a major oil spill in the United States is very high and, perhaps as a consequence, there have been no extremely large spills in the United States since the Exxon Valdez oil spill. There have been spills that might have become very large and caused widespread injuries if it had not been for the preplanned aggressive response effort undertaken.[44] This lack of extremely large oil spills in the United States for over a decade has had an interesting effect; it implies that while CV has not been used much for assessing natural resource damage of large oil spills, its potential use may be playing an important role in preventing such spills.[45] Elsewhere, the pattern of big oil spills has been largely unchanged (Chapple 2000).

Acknowledgements

The authors wish to acknowledge the many contributions provided by Michael Conaway and Kerry Martin during the course of this study. Richard Bishop, Gardner Brown, Howard Schuman, Norbert Schwarz, Paul Slovic, and Robert Solow provided comments during the course of the study. We also thank Jon Krosnick, V. Kerry Smith, Jon Strand and *journal* referees who provided comments on earlier drafts. The State of Alaska provided funding for this study. Kopp acknowledges the partial support of the Alfred P. Sloan Foundation to the Welfare Economics Program at Resources for the Future. All opinions expressed in this paper are those of the authors and should not be attributed to the State of Alaska, the Alfred P. Sloan Foundation, or the authors' home institutions. The authors bear sole responsibility for any errors or omissions.

Notes

1. Descriptions of the grounding of the Exxon Valdez may be found in National Safety Transportation Board (1990) and Moore (1994). A number of spill prevention and containment measures were put into place when oil first began to be shipped from Valdez. These measures were intended to reduce various types of risks that had been identified in an initial comprehensive risk assessment. That assessment had identified one of its most likely bad accidents as a tank hitting the reef next to Bligh Reef under somewhat similar conditions as the Exxon Valdez. These measures had been progressively "relaxed" over time, in part due to their expense and in part because there had not previously been any really serious accident. These measures might have prevented or largely contained the Exxon Valdez spill had they been in place at the time of the spill.
2. *Ohio v. Department of Interior*, 880 F.2d 432 (D.C. Cir. 1989).
3. See *Ohio v. Department of Interior*, 1989. The original Department of the Interior rules challenged in the *Ohio v. DOI* case were published in the *Federal Register*, vol. 51, August 1, 1986. See Kopp, Portney and Smith (1990) for a comprehensive discussion of the *Ohio* decision.
4. Following the *Ohio* decision, the U.S. District Court of Utah in a 1992 CERCLA case rejected a proposed consent decree, in part, for failing to include lost passive use values in the determination of damage associated with groundwater contamination. See *State of Utah v. Kennecott Corporation*, No. CIV 86-0902G, United States District Court, D. Utah, September 3, 1992, Memorandum Decision and Order.
5. This position is consistent with OPA legislative history that specifically refers to diminution in value as a part of damages and cites the *Ohio* decision definition of value, which includes both direct use and passive use.
6. These models and methods are termed indirect approaches and include the travel cost model and the hedonic property value model. An introduction to the use of these models for the assessment of damages due to natural resource injuries can be found in McConnell (1993).
7. A comprehensive discussion of contingent valuation is contained in Mitchell and Carson (1989).
8. For a sense of the debate immediately post Exxon Valdez, see the 1994 *Journal of Economic Perspective* Symposium papers by Diamond and Hausman, Hanemann, and Portney. For a more recent review see Carson, Flores and Meade (2001).
9. Due to space limitations, many details of the study and the complete survey instrument could not be incorporated into this paper. The complete text of the report can be found online at Exxon Valdez Oil Spill Trustee website: http://www.oilspill.state.ak.us/gem/facts/econ omic.html. The complete survey instrument including color copies of visual material used can be found at:

http://www.econ.ucsd.edu/~rcarson as can the complete dataset from the study. Mitchell and Carson (1995) and Mitchell (2002) provide additional discussion of various issues involved in the design of contingent valuation surveys.

10. The description of the injuries was based on scientific information provided to the study team by the State of Alaska. There was substantial uncertainty regarding the precise extent of some of the injuries at the time the final survey was conducted. In order to minimize the litigation risk associated with that uncertainty, the study team valued a conservative representation of the injuries. Therefore, only injury facts of which scientists were reasonably certain as of the fall of 1990 were used. When the best estimate of the actual state of affairs required a range, the conservative end of that range was used.
11. Willingness to accept is the appropriate property right for natural resource damages. Respondents in CV surveys tend to find questions that ask them how much they would accept in compensation to voluntarily accept a loss to a public environmental good implausible for a variety of reasons (Mitchell and Carson 1989), since they do not believe they possess a personal property right to sell the good.
12. The identity of the survey's sponsor, the State of Alaska, was not revealed to either the interviewers or the respondents.
13. This type of qualitative research is increasingly used by survey researchers in the early stages of designing contingent valuation questionnaires because they are an efficient way to explore people's beliefs, attitudes, and knowledge about the good to be valued, and to obtain their reactions to possible CV scenario elements (Morgan 1993).
14. In later groups, elements of a possible questionnaire were described in more detail to help us understand how the participants understood these elements and how they used this information. These included the payment vehicle, duration of payments, description of the injuries, description of a plan to prevent future spills, and use of particular photographs and maps to communicate factual aspects of the scenario.
15. Westat, one of the country's premier survey research organizations conducted the interviews for this study, recruited the professional interviewers (who gave face-to-face interviews at the respondent's home), prepared the interview materials based on the instrument we delivered to them, conducted the interviewer training, supervised the production of interviews in the field, and edited and validated the completed questionnaires.
16. Pilot I reflects the first formal field test. Pilot II, a split-sample test, compared the effect of two possible payment vehicles, income taxes and oil prices. Pilot III encompassed two split-sample tests comparing: a) revised versions of the income tax and oil price payment vehicles and, b) effect of excluding an environmental item in each of questions A-1 and A-3 on the WTP estimate. Pilot IV was the last formal field test and closely resembled the final survey.
17. This downward bias is suggested by empirical evidence and probably results from expectations formed by the initial cost estimate given to the respondent. Some respondents who vote to pay the first amount might be willing to pay the second (higher) amount but vote against the higher amount when asked because they feel that the government would waste the extra money requested. In addition, some respondents who are not willing to pay the first amount would be willing to pay the second (lower) amount but may vote against the second amount because they believe that either the government will deliver a lower quality good than that first promised or that the probability of the government delivering the good is lower at the lower price. Both of these voting patterns result in a downward bias. The extent of the bias depends on the degree to which the second amount is perceived by the respondent as being an independent cost estimate. Carson, Groves and Machina (1999) provide a formal conceptual framework for considering this issue.
18. The telephone survey valued the installation of a scrubber on a power plant in Columbus, Ohio using 500 observations in a split-sample design with a lump sum and annual payment schemes. While the results from this survey clearly rejected ($p < 0.01$) Kahneman and Knetsch's (1992,

p. 63) contention that respondents do not focus on the temporal nature of the payment obligation, they were consistent with the presence of high discount rates and/or borrowing constraints observed with many consumer durables.

19. Embedding is a term introduced into the contingent valuation literature by Kahneman to refer to various issues related to the sequencing and nesting of goods as well as a survey design problem known as part-whole bias. See Carson and Mitchell (1995) for a discussion.
20. For discussions, see Hoehn and Randall (1989); Bishop and Welsh (1992); and Carson, Flores and Hanemann (1998).
21. Potential respondents were told that the interview was for a study of people's views about current issues.
22. A line drawing of an escort ship recovering oil at an oil spill proved to be very helpful in explaining how the escort program would work.
23. Pretests had shown that some respondents criticized the notion that citizens should share in paying the cost of the plan. Because this could lead respondents to reject the premise of the scenario, that they should make a judgment about what the plan is worth to them, the interviewers were instructed to say the following to those who expressed the view that Exxon or the oil companies should pay in an attempt to persuade them that the oil companies *would* pay a share: "If the program is approved, the oil companies that bring oil through the Alaska pipeline (including Exxon) *will* have to pay part of the cost by a special tax on their corporate profits."
24. The first pilot study established a large fraction of the population was willing to pay small amounts for the program while zero percent was willing to pay $1000. After this effort was devoted to helping to get reasonable estimates of the fraction in favor at more central quantiles of the WTP distribution.
25. Shortly after completing each interview, the interviewer completed a series of questions in Section D about the circumstances under which the interview was conducted and the interviewer's impressions about whether the respondent had any difficulty understanding the vote questions and the seriousness of the consideration the respondent gave to these questions.
26. Before the selection was made, the 1,179 PSU's were stratified by the following 1980 Decennial Census characteristics: (1) region of the country; (2) SMSA versus non-SMSA; (3) rate of population change between 1970 and 1980; (4) percent living on a farm (for non-SMSA PSU's); (5) percent employed in manufacturing; (6) percent white; (7) percent urban; and (8) percent over age 65. Selection from strata typically increases the precision of the survey results compared to unstratified selection. For a discussion of the comparative advantages of stratified selection, see Sudman (1976). The 1980 census was used for the sample, as results from the 1990 census were not yet available.
27. Due primarily to logistical and cost considerations, no foreign language versions of the questionnaire were developed. As a result, non-English speaking households were not eligible to be interviewed. Thus, we correspondingly reduced the 1990 Census estimate of the number of U.S. households (93,347,000) by 2.7%, our survey's estimate of the proportion of U.S. households that were non-English speaking. This yields a population of 90,838,000 English-speaking households to which our results may be extrapolated.
28. For most of the respondents giving "not sure" answers, this interpretation seems to be appropriate. Some respondents gave a "not sure" answer to A-15 and subsequently gave a "yes" answer to the substantially lower amount in A-17. Similarly, some respondents gave "yes" responses to A-15 and "not sure" responses to the higher amount in A-16. A likely interpretation is that these "not sure" responses represent respondents who were reasonably close to their indifference thresholds. Of the 141 respondents who gave one or more "not sure" responses, 111 followed this pattern. The other 30 respondents (less than 3% of the sample) gave "not sure" responses to both A-15 and A-17; these respondents may not have been capable of answering the WTP questions. We have also conservatively treated these as no-no responses.

29. In the analysis that follows we have also assumed that respondents do not engage in non-truthful preference revelation with respect to the second choice question. The strong theoretical prediction (Carson, Groves and Machina 1999) is that the response to the second question should be inconsistent with the first. Under stronger but plausible conditions, the response to the second question will be consistent with lower willingness to pay than the first question. The empirical results obtained here are consistent with this prediction in that an analysis based upon only the first question results in a larger WTP estimate than using both questions. Useful information can be obtained from a second question without assuming consistency between questions, but the statistical modeling is much more complex and more dependent upon assumptions made (Alberini, Kanninen and Carson 1997).
30. If the amenity being valued is a "bad" to the respondent, then the lower bound on the interval is potentially negative infinity rather than zero. While a possibility with some public goods, it is unlikely that anyone views an Exxon Valdez type oil spill as something desirable.
31. The WTP intervals of the 10 respondents who indicated that they wanted to change their votes were set from zero to the highest amount to which they had previously said they would vote "for." In addition, four respondents who did not answer the second WTP question (A-16 or A-17) had their WTP intervals based only on their response to A-15.
32. The Weibull is the simplest distribution that allows an increasing, decreasing, or constant hazard function. It is also flexible enough to approximate several other commonly used survival distributions, such as the exponential, the Raleigh, the normal, and the smallest extreme value.
33. From this point on we use the household weights provided by Westat in performing any estimations. The differences between the weighted and unweighted estimates are almost always quite small, with the weighted estimates being slightly lower than the unweighted estimates. The construction of the weights is discussed at length in section 4.10 and Appendix B.3 of the original study and are based solely on Census demographic variables. Unweighted and weighted frequencies are provided in Appendix C.1.
34. Since this estimate is a linear function of predetermined design points and a multinomial variable, a standard error for this estimate $2.71, and a 95% confidence interval [$48.28–$58.91] are straightforward to calculate. The Turnbull lower bound on the mean increases toward the mean from below as more design points are added. However, for a fixed sample size adding more design points causes the variance of this estimator to grow and hence this estimator can be seen to represent the commonly found bias-variance tradeoff.
35. It is possible to fit a number of other common two-parameter survival distributions to our data. These tend to result in similar estimates of the median WTP but quite different estimates of the mean. The fit of the Weibull distribution is either statistically superior or indistinguishable with respect to these other distributions. More flexible three parameter distributions tend to suggest a sharper drop-off in the percent willing to pay near zero and a sharper drop-off in the right tail. As a result, estimates of mean WTP from these distributions tend to fall between the lower-bound estimate from the Turnbull estimator and that of the two parameter Weibull.
36. The log-likelihood for the Weibull spike model (Kristrom 1997) is –1331.293 so a likelihood ratio test rejects the two parameter Weibull model in favor of the Weibull spike model at $p < 0.01$. Effectively what is happening is that the better fit near zero reduces the implied variance of the WTP distribution. This reduction in variance (reflected in a decreased scale parameter) pulls in the right tail of the distribution, and hence, reduces the estimate of mean WTP.
37. Because most of the missing values are on income, we have estimated an equation to predict the log of income. The estimated coefficients for this equation, which is based largely on demographic characteristics, are provided in the study report. All of the variables have the expected sign, and the equation has an R^2 of 0.46.
38. The Turnbull lower bound on the mean should approach the true mean from below as one increases the number of design points. The main reason for not using a large number of design points in conjunction with the Turnbull approach is that the variance of the estimate can increase

rapidly as a fixed sample size is randomly allocated to a larger and larger number of design points. While the Turnbull estimates of the lower bound on the mean from the single and double bounded data from this study are potentially consistent with each other if the actual WTP distribution closely follows a step function, this seems unlikely. A more plausible explanation for the similarity between the two estimates is that the downward bias induced in the second question just offsets the increase in the lower bound estimate of the mean that would be found from increasing the number of design points in the double bounded framework. As such, some of the statistical gain from using the double bounded approach may be illusory, although the information from it may be useful in a more general context.

39. The damage assessment regulations under OPA also note that WTA is the appropriate measure of damages: "Because the government is holding natural resources in trust for the public, the WTA criterion is conceptually the more appropriate measure of damages for natural resource damage claims." *Federal Register*, vol. 59, January 7, 1994, p. 1150.
40. This number is obtained by multiplying the Turnbull lower bound mean by the number of English-speaking U.S. households (90,838,000).
41. There were also a number of private claims for commercial and punitive damages brought by private parties. While many of those cases have been settled, some litigation continues as of 2003.
42. Much has also been learned about structuring restoration and resource compensation plans. For details in this case, see http://www.oilspill.state.ak.us/.
43. The State of California commissioned an *ex ante* study (Carson et al., forthcoming) of the benefits of preventing oil spills along California's central coast and has used it for a number of different policy purposes. This study builds upon the Exxon Valdez study reported in this paper. It incorporates a number of refinements to that survey and was intentionally designed so that its survey instrument could be more readily adaptable to other geographic areas.
44. Most oil spill injuries in the U.S. since the Exxon Valdez have been short term to outdoor recreation or to small parts of larger ecosystems for which the government is the trustee. Stated preference techniques, indirect techniques like travel cost analysis, and habitat equivalence analysis have been used to help settle these cases. These spills have also caused some harm to commercial interests who can bring private lawsuits where lost profits, current or perspective, are at issues.
45. The vast majority of CV studies have always been done for policy purpose. The number of such studies continues to grow rapidly. Carson (forthcoming) provides citations to over 5000 CV papers and studies from over 100 countries.

References

Alberini, Anna and Richard T. Carson (1993), 'Choice of Thresholds for Efficient Binary Discrete Choice Estimation'. Discussion Paper 90-34R, Department of Economics, University of California, San Diego.

Alberini, A., B. Kanninen and R.T. Carson (1997), 'Modeling Response Incentives in Dichotomous Choice Contingent Valuation Data', *Land Economics* **73**, 309–324.

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* **58**, 4601–4614.

Bateman, Ian, Richard T. Carson, Brett Day, W. Michael Hanemann, Nick Hanley, Tannis Hett, Michael Jones-Lee, Graham Loomes, Susana Mourato, Ece Özdemiroglu, David Pearce, Robert Sugden and John Swanson (2002), *Economic Valuation with Stated Preference Techniques: A Manual*. Northhampton, MA: Edward Elgar.

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 Michael P. Welsh (1992), 'Existence Values in Benefit-Cost Analysis and Damage Assessment', *Land Economics* **68**, 405–417.

Carson, Richard T. (1997), 'Contingent Valuation Surveys and Tests of Insensitivity to Scope', in R.J. Kopp, W. Pommerhene and N. Schwartz (eds.), *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*. Boston: Kluwer.

Carson, Richard T. (forthcoming), *Contingent Valuation: A Comprehensive Bibliography and History*. Northampton, MA: Edward Elgar.

Carson, Richad T., Michael B. Conaway, W. Michael Hanemann, Jon A. Krosnick, Robert Cameron Mitchell and Stanley Presser), *Valuing Oil Spill Prevention: A Case Study of California's Central Coast*. Boston: Kluwer Academic Press, forthcoming.

Carson, Richard T., Nicholas E. Flores and W. Michael Hanemann (1998), 'Sequencing and Valuing Public Goods', *Journal of Environmental Economics and Management* **36**, 314–323.

Carson, Richard T., Nicholas E. Flores and Norman F. Meade (2001), 'Contingent Valuation: Controversies and Evidence', *Environmental and Resource Economics* **19**, 173–210.

Carson, Richard T., Nicholas E. Flores and Robert C. Mitchell (1999), 'The Theory and Measurement of Passive Use Value', in I.J. Bateman and K.G. Willis (eds.), *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EC, and Developing Countries*. Oxford: Oxford University Press.

Carson, R.T., T. Groves and M. Machina (1999), 'Incentive and Informational Properties of Preferences Questions', Plenary Address, European Association of Environmental and Resource Economists, Oslo, Norway.

Carson, Richard T., W. Michael Hanemann, Raymond J. Kopp, Jon A. Krosnick, Robert C. Mitchell, Stanley Presser, Paul A. Ruud and V. Kerry Smith (1997), 'Temporal Reliability of Estimates from Contingent Valuation', *Land Economics* **73**, 151–163.

Carson, Richard T. and Robert Cameron Mitchell (1995), 'Sequencing and Nesting in Contingent Valuation Surveys', *Journal of Environmental Economics and Management* **28**, 155–173.

Carson, Richard T. and Dan Steinberg (1990), 'Experimental Design for Discrete Choice Voter Preference Surveys', in *1989 Proceeding of the Survey Methodology Section of the American Statistical Association*. Washington: American Statistical Association.

Carson, Richard T., Leanne Wilks and David Imber (1994), 'Valuing the Preservation of Australia's Kakadu Conservation Zone', *Oxford Economic Papers* **46**(S), 727–749.

Chapple, Clive (2000), 'The 1990 Oil Pollution Act: Consequences for the Environment', paper presented at the Association of Environmental and Resource Economists summer workshop, La Jolla, CA.

Diamond, Peter and Jerry A. Hausman (1994), 'Contingent Valuation: Is Some Number Better Than No Number', *Journal of Economic Perspectives* **8**(4), 45–64.

Fararo, Kim (1992), 'Near Miss in the Narrows Oil-Laden Tanker's Tugboat Escort Puts Itself in a Hard Spot Between Middle Rock and a New Disaster', *Anchorage Daily News*, November 22, A1.

Flores, Nicholas E. and Jennifer Thatcher (2002), 'Money Who Needs It?: Natural Resource Damage Assessment', *Contemporary Economics Policy* **20**, 171–178.

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

Hanemann, W. Michael (1991), 'Willingness to Pay and Willingness to Accept: How Much Can They Differ?' *American Economic Review* **81**, 635–647.

Hanemann, W. Michael (1994), 'Valuing the Environment Through Contingent Valuation', *Journal of Economic Perspectives* **8**(4), 19–43.

Hanemann (1999), 'Neo-Classical Economic Theory and Contingent Valuation', in I.J. Bateman and K.G. Willis (eds.), *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EC, and Developing Countries*. Oxford: Oxford University Press.

Hanemann, W. 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, Jerry A. (ed.) (1993), *Contingent valuation: A Critical Assessment*. Amsterdam: North-Holland.

Hoehn, John P. and Alan Randall (1989), 'Too Many Proposals Pass the Benefit Cost Test', *American Economic Review* **79**, 544–551.

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

Kopp, Raymond J., Paul R. Portney and V. Kerry Smith (1990), 'The Economics of Natural Resource Damages After *Ohio v. U.S. Department of the Interior*', *Environmental Law Reporter* **20**(4), 10127–10131.

Kristrom, Bengt (1997), 'Spike Models in Contingent Valuation', *American Journal of Agricultural Economics* **79**, 1013–1023.

Krueger, Richard A. (1988), *Focus groups: A Practical Guide for Applied Research*. Newbury Park, CA: Sage Publications.

Louviere, Jordan J., David A. Hensher and Joffre D. Swait (2000), *Stated Choice Methods: Analysis and Applications*. New York: Cambridge University Press.

McConnell, Kenneth E. (1993), 'Indirect Methods for Assessing Natural Resource Damages Under CERCLA', in R.J. Kopp and V.K. Smith (eds.), *Valuing Natural Assets: The Economics of Natural Resource Damage Assessment*. Washington: Resources for the Future.

Mitchell, Robert Cameron (2002), 'On Designing Constructed Markets in Valuation Surveys', *Environmental and Resource Economics* **22**, 297–321.

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

Mitchell, Robert Cameron (1995), 'Current Issues in the Design, Administration, and Analysis of Contingent Valuation Surveys', in P.O. Johansson, B. Kristrom and K.G. Mäler (eds.), *Current Issues in Environmental Economics*. Manchester: Manchester University Press.

Moore, William H. (1994), 'The Grounding of the Exxon Valdez: An Examination of the Human and Organizational Factors', *Marine Technology* **31**, 41–51.

Morgan, David L. (ed.) (1993), *Successful Focus Groups: Advancing the State of the Art*. Newbury Park, CA: Sage Publications.

National Transportation Safety Board (1990), *Grounding of the U.S. Tankership Exxon Valdez on Bligh Reef, Prince William Sound Near Valdez Alaska*, PB90-916405. Washington: National Safety Transportation Board.

Nelson, Wayne (1982), *Applied Life Analysis*. New York: John Wiley.

Portney, Paul R. (1994), 'The Contingent Valuation Debate: Why Economists Should Care', *Journal of Economic Perspectives* **8**(4), 3–17.

Sudman, Seymour (1976), *Applied Sampling*. New York: Academic.

Turnbull, Bruce W. (1976), 'The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data', *Journal of the Royal Statistical Society, Series B* **38**, 290–295.