DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

THE USE AND MISUSE OF SURVEYS IN ECONOMIC ANALYSIS: NATURAL RESOURCE DAMAGE ASSESSMENT UNDER CERCLA

Charles J. Cicchetti
Putnam, Hayes & Bartlett, Inc.

Jeffrey A. Dubin
California Institute of Technology

Louis L. Wilde
California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 768

July 1991

THE USE AND MISUSE OF SURVEYS IN ECONOMIC ANALYSIS: NATURAL RESOURCE DAMAGE ASSESSMENT UNDER CERCLA

Charles J. Cicchetti, Jeffrey A. Dubin and Louis L. Wilde

ABSTRACT. This paper examines problems with the admissibility of contingent use methodology surveys in natural resource damage assessment cases under the Comprehensive Environmental Response, Compensation and Liability Act of 1980 (CERCLA), as well as the propriety of their use in formulating public policy. Using a contingent use survey conducted in conjunction with the New Bedford Harbor Superfund case and two follow-up surveys, a number of errors and biases associated with contingent use methodology surveys are isolated and analyzed.

THE USE AND MISUSE OF SURVEYS IN ECONOMIC ANALYSIS: NATURAL RESOURCE DAMAGE ASSESSMENT UNDER CERCLA

Charles J. Cicchetti, Jeffrey A. Dubin and Louis L. Wilde[1]

1 Introduction

There are two basic questions concerning the use of surveys in litigation: are the surveys admissible as evidence; and if they are not, under Rule 703 of the Federal Rules of Evidence (Green and Nesson 1988), may surveys be relied upon by an expert in his or her trial testimony. These questions are important given the inherent hearsay character of survey evidence; survey respondents normally are unavailable for cross-examination. The courts have addressed these questions by establishing various "foundational criteria" which must be satisfied in order for survey evidence to be admissible:

- the persons conducting the survey must be recognized experts;
- the data gathered must be accurately reported;
- the sample design, the questionnaire and the interviewing must be in accordance with generally-accepted standards of objective procedure and statistics in the field of such surveys, which include proper definition of the "universe," selection of a representative sample of that universe and framing of questions in a clear, precise and nonleading manner;
- the sample design and interviews must be conducted independently of the attorneys; and
- the interviewees and the interviewers (trained in this field) must have no knowledge of the litigation or the purpose for which the survey is to be used.[2]

---

[1] Charles J. Cicchetti is Co-Chairman of Putnam, Hayes & Bartlett, Inc. Jeffrey A. Dubin and Louis Wilde are Associate Professor and Professor of Economics at the California Institute of Technology. The authors thank Ellen K. Moran, Kristina M. Sepetys and Mary-Elizabeth Vault for their contributions to the paper.

[2] For further discussion, see Cicchetti and Peck (1989). These criteria were originally promulgated by the Judicial Conference of the United States and have been affirmed by the courts. See, *e.g.*, *Toys "R" Us, Inc. v. Canarsie Kiddie Shop, Inc.*, 559 F. Supp. 1189, 1205 (E.D.N.Y. 1983); *Bank of Utah v. Commercial Security Bank*, 369 F.2d 19 (10th Cir. 1966), cert. denied, 386 U.S. 1018 (1967); Manual for Complex Litigation

While surveys satisfying these foundational criteria have been admitted as evidence in some cases (*e.g.*, to establish consumer confusion in trademark infringement cases and to establish markets in antitrust cases),[3] they have only begun to be admitted to establish damages. In a 1989 Circuit Court decision, a particular survey technique known as the *contingent valuation method* (CVM) was upheld under the Comprehensive Environmental, Response Compensation Liability Act (CERCLA) as "the methodology qualified as best available procedure for determining damages flowing from destruction of or injury to natural resources."[4] The CVM surveys individuals to determine their willingness to pay for environmental goods or services under various hypothetical circumstances. This data is then used to calculate damage estimates.

The *contingent use method* (CUM) is another damage assessment technique that relies on survey data. The CUM uses survey trip data, typically combined with travel cost data, to estimate demand curves for environmental goods or services before and after a specified improvement or injury. These demand curves are then used to calculate nonmarket values. Contingent use analyses are similar to travel cost analyses in overall methodological structure, but different with regard to data source. The CUM relies on hypothetical survey data, whereas the travel cost method relies on non-hypothetical survey data or actual observed data. Although slightly different from

---

116 (5th ed. 1982); *Handbook of Recommended Procedures for the Trial of Protracted Cases*, 25 F.R.D 351, 429 (1960); *Pittsburgh Press Club v. United States*, 579 F.2nd 751, 758 (3d Cir. 1978); *Wuv's International, Inc. v. Love's Enterprises, Inc.*, 208 U.S.P.Q. 736, 754 (D. Colo. 1980); *Saiba v. State*, 475 N.E.2d 1181, 1187-88 (Ind. App. 1985).

[3] Early trademark infringement cases include *Oakite Products, Inc. v. Buckeye Soda Co.* (1930) 6 US Pat Quart 1952, 19 Cust & Pat App. (Pat 1034, 56 F2d 462) and *Procter & Gamble v. J. L. Prescott Co.* (1939, CA3 NJ) 102 F2nd 773, cert den 308 US 557, 84 Led 468, 60 S Ct. 80. A more recent example in *Envirosafe Services, Inc. v. Envirosure Management Corporation* (1989), No. 87-4659, Slip Op. (E.D.Pa., January 5). Antitrust cases include *United States v. J. I. Case Co.* (1951, DC Minn) 101 F Supp 856 and *United States v. E. I. DuPont De Nemown & Co.* (1959, DC 111) 177 F Supp 1. Morgan (1990) provides a systematic review of cases in which surveys or opinion polls were accepted into evidence.

[4] *Ohio v. United States Department of Interior*, 880 F.2nd 432 (D.C. Cir. 1989) 436. CERCLA, also known as Superfund, permits the United States government or any state as public trustee of natural resources, to sue certain parties for damages "for injury to, destruction of, or loss of natural resources" caused by hazardous waste disposal. 42 U.S.C. 9601 *et seq.*. CERCLA defines the term "natural resources" broadly to include "land, fish, wildlife, biota, air, water, ground water, drinking water supplies and other such resources," 42 U.S.C. 9601(16) (ELR 44006). As of 1989, DOI had applied the Natural Resource Damage Assessment Regulation (43 U.S.C. Part 11) only once. In that case, an oil tanker spilled 500,000 gallons into the Savannah River after a valve malfunctioned. DOI calculated damages based largely on lost opportunities for hunting and fishing in an adjacent wildlife refuge. The owner of the tanker paid over $1.2 million. (John Lancaster, "Method for Assessing Oil Spill Damages Hit Environmentalist Fault Interior," Washington Post 26 June 1989, 1.)

contingent valuation in that it does not elicit a value, contingent usage has been relied upon to estimate the lost use value of non-market resources and services.

General discussion of the CVM and its potential problems are commonplace,[5] but few problems have been documented in case studies. One example is provided by Randall and Stoll (1983). They describe a study by Schulze, Brookshire and Thayer (1981) which estimated "the annual value to Chicago residents of one particular increment in visibility at the Grand Canyon at $86 for a typical household." They then describe a subsequent study by Randall, Hoehn and Tolley (1981), also based on Chicago residents, in which

> [s]tarting with the immediate Chicago region, the visibility increment was valued at about $325 per household annually. When the region was expanded to include all of the U.S. east of the Mississippi, the program was valued at about $355. When a visibility improvement program for the Grand Canyon was added, the whole package was valued at about $373. The incremental value of the Grand Canyon program was $18, compared to $86 when that program was considered alone.[6]

As the CVM becomes the subject of significant study and controversy, there will likely be increased reliance on the CUM, since the CUM only requires individuals to predict usage levels, not actual values. However, although individuals may find it easier to estimate usage rather than value, the CUM has problems of its own. Cicchetti and Peck (1989), for example, discuss several practical issues associated with the use of the CVM and the CUM in litigation, as well as the propriety of using these techniques for public policy.

This paper will document some of the problems with the CUM in the context of an actual case study. In particular, using an initial survey and two follow-up surveys developed for the New Bedford Harbor Superfund case between March 1986 and

---

[5] See Mitchell and Carson (1989) and Cummings, Brookshire and Schulze (1986) for critiques of the CVM as well as responses to those critiques.

[6] Majid, Sinden and Randall (1983) similarly found that the estimated values of proposed parks were greater when the proposed parks were considered in isolation rather than as additions to an existing system of parks. On the other hand, Dickie, Fisher and Gerking (1987) found little difference between the estimated demand curves for strawberries based on actual market transaction data and hypothetical responses. The latter result supports the contention that CVM works best for familiar goods. In this case a well-defined market for strawberries exists, so it should not be too surprising if individuals' hypothetical responses mimic the market.

September 1987 we isolate and analyze several specific errors and biases associated with the CUM.

A number of interesting conclusions regarding the reliability of CUM surveys and their role in CERCLA litigation can be drawn from our analysis. For example, we find that awareness of a particular pollutant does not explain whether an individual predicts that he or she would use a beach more absent that pollutant. Furthermore, most individuals who claim that they would use a beach more if the pollutant were cleaned up, change their responses when reminded that other pollutants will remain. In this case, those individuals who are more aware of the pollutant are more likely to change their responses.

Aside from the general issue of survey reliability, one important question is whether, given the current state of development of contingent techniques does an adequate consensus exist among economists about the proper application of contingent techniques to permit them to be admitted in court? As the following analysis suggests, in the New Bedford Harbor case at least, it is questionable whether the initial survey would have met the fundamental criteria which must be satisfied in order for survey evidence to be admissible.[7]

The paper is organized as follows. Section 2 provides an overview of the study of response error in surveys, including a brief discussion of the CVM. Section 3 describes the New Bedford Harbor surveys that are the subject of this paper. Section 4 analyzes specific errors and biases associated with the surveys, and Section 5 contains concluding comments on the policy implications of the analysis.

2 Historical Use of Surveys

Surveys have long been used in a variety of disciplines, including psychology, political science and economics. But many economists are skeptical about the usefulness of survey data because of its susceptibility to a variety of errors and biases.

[7] A recent example of a problematic CVM analysis is provided by *State of Idaho v. Southern Refrigerated Transport, Inc.* (D. C. Idaho 1991). In this case, the State of Idaho attempted to determine the value of a non-returning fish population by a contingent valuation survey. The survey was administered to determine what individuals in the Northwest would be willing to pay in the form of increases to their power bills to double the runs of two types of fish in an entire river. In this case, the Court found that the study was "not pervasive and it would be conjecture and speculation to allow damages based on [it]."

Not surprisingly, survey errors and biases have been of concern to others besides economists. In fact, systematic studies of response error in surveys began in the early 1950s, at roughly the same time that modern survey techniques were being developed.

In an early contribution to the survey literature, Parry and Crossley (1954) analyzed responses to a variety of factual questions which could be verified by public records. Invalid answers ranged from nearly zero for questions regarding telephone ownership to one-third and more for voting and contributions to Community Chest.[8] Kish and Lansing (1954) compared homeowners' estimates of the market value of their houses in the 1950 Survey of Consumer Finances to estimates for the same houses made by professional appraisers. While the mean of homeowners' estimates was not statistically significantly different than the mean of appraisers' estimates, only 37 percent of the homeowners' estimates were within plus or minus 10 percent of the appraisers' estimates.[9] More recently, Chase and Godbey (1983) analyzed the accuracy of self-reported participation rates for members of a tennis club and a swimming club. Both groups significantly overestimated their participation rates; 56 percent of the tennis club members by more than 100 percent, and 43 percent of the swimming club members by more than 100 percent.[10]

The crime literature and the marketing literature also have been concerned with the quality of survey recall data, as well as the quality of hypothetical survey data in the case of the marketing literature. For example, Wyner (1980) examined the accuracy of self-reported lifetime arrests for ex-heroin addicts in New York City. While the mean of the actual number of arrests was virtually identical to the mean of the self-reported number of arrests (9.25 versus 9.6), only 10 of 79 respondents were completely accurate and more than 20 percent made errors of plus or minus five or more

---

[8] Items of investigation in the Parry and Crossley study were (1) registration and voting in city-wide Denver elections, (2) contributions to Community Chest, (3) possession of a Denver Public Library Card, (4) possession of a Colorado driver's license, (5) automobile ownership, (6) age, (7) homeownership and (8) telephone ownership.

[9] In another study, Kain and Quigley (1972), reporting on a 1967 survey, obtained similar results. In their study, however, only 28 percent of the owners of single detached structures provided estimates of value within 10 percent of appraisers' estimates of value for the same homes.

[10] See Chase and Harada (1984) for a detailed analysis of response errors in the Chase and Godbey survey.

arrests.[11] Juster (1966), in a study of purchase intentions and actual purchase behavior, found that only 53 percent of those who said they were certain to buy a new automobile, and only 21.5 percent of those who said they were certain to buy a new major appliance, actually did so.[12] Newman and Lockeman (1975) compared survey-based and observation-based measures of pre-purchase information seeking. Their study "found considerable search activity in the retail store and showed it to be understated by the usual survey-based counts."

Within the economics literature, there has been a very recent effort to analyze the limits of the CVM. A relatively large literature compares the CVM to travel cost or hedonic methods.[13] Kealy, Dovidio and Rockel (1988) and Kealy, Montgomery and Dovidio (1990) use "simulated markets" to test various aspects of the CVM related to reliability and predictive validity.[14] Loomis (1989) analyzes the test-retest reliability of the CVM. Reiling, Boyle, Phillips and Anderson (1990) examine one aspect of the temporal reliability of the CVM by asking whether estimated contingent values "vary according to the time period when a study is conducted." Finally, Huppert (1989) points to the similarities in the results from the travel cost method and the CVM when comparing willingness-to-pay estimates. Overall, these studies are inconclusive; depending on one's point of view, they support either the claim that, where comparable, the CVM is as accurate as any other method for estimating damages or the claim that the CVM is too unreliable to be of practical use.

---

[11] Miller and Groves (1985) discuss generally the use in the crime literature of external record evidence to evaluate survey responses.

[12] See also Granbois and Summers (1975) on the predictive accuracy of "intention measures," and Morrison (1979) for further analysis of Juster's data.

[13] See, for example, Bishop and Heberlein (1979); and Smith, Desvousges and Fisher (1986).

[14] The CVM literature makes a distinction between validity and reliability. Validity studies are designed to ascertain whether the correct theoretical construct is being measured and whether the statistically estimated mean is equal to the true mean for the item being evaluated. (Reiling, Boyle, Phillips and Anderson 1990, 128). Reliability, on the other hand, is the extent to which the variance of the item being measured in a survey is due to random sources. For further discussion see Mitchell and Carson 1989, 189-229.

3 The New Bedford Harbor Survey Instruments

The surveys analyzed in this article grew out of a CERCLA case in New Bedford, Massachusetts.[15] Plaintiffs (the federal government) alleged that they had suffered injury and damages to recreational resources in the New Bedford Harbor area as a result of alleged polychlorinated biphenyl (PCB) contamination. A survey was undertaken by the federal government to determine whether, and to what extent (if at all), beach use in the New Bedford Harbor area had been affected. This survey was designed for the government by Industrial Economics, Inc. (IEC) under the direction of Dr. Kenneth E. McConnell and conducted in March, 1986, by Decision Resource Corporation (DRC), and is referred to herein as "Survey 1." The other two surveys were designed, administered and conducted under the joint direction of Drs. Charles J. Cicchetti and Bernie Reddy in May and September of 1987 by Mathematica Policy Research, Inc. (MPR). They are referred to herein as "Survey 2" and "Survey 3," respectively.

3.1 Survey 1

Telephone calls were made to 545 residents of the towns of New Bedford, Fairhaven and Dartmouth, drawn from a sample of 2,000 households. The respondents were screened to establish eligibility using questions regarding age, residency and status as a decision-maker regarding household beach attendance.

All eligible respondents were asked whether they or anyone in their household had visited any beaches in the Fairhaven-New Bedford-Dartmouth area in 1985 and, if so, which ones (there are 14 public and private beaches in the area). For six of the beaches which could have been mentioned, follow-on questions were asked regarding actual number of visits during 1985 (typical length of stay, usual mode of transportation, approximate travel time and the number of planned visits for 1986).

---

[15] In 1983 the Department of Justice (DOJ) filed a civil complaint against five companies, charging them with responsibility for releasing PCBs into the Harbor. An agreement in principle was reached in February 1990 with two defendants, Aerovox, Inc. and Belleville Industries Inc. In September 1990, AVX and EPA agreed to a $66 million settlement, one of the largest settlements by a single defendant in the 10-year history of Superfund. The settlement money will fund the clean-up and reimburse funds already spent in a variety of enforcement, remedial investigation and damage assessment activities. The remaining two defendants, Delaware-based FPE and Cornell-Dubilier, have also reached a settlement in principle.

All respondents, both the beach-goers and the non-beach-goers, were asked to rate on a scale the environmental quality of the New Bedford Harbor. They were then asked to name specific substances or chemicals (if any) that they thought were damaging the Harbor. Those respondents who were not aware of any substances, or who identified substances other than PCBs, were then asked if they believed that the Harbor is contaminated with PCBs.

Respondents who were "aware" of PCBs (defined as both those who volunteered PCBs and those who responded affirmatively to the direct question about the presence of PCBs) were asked when (in what year) and how they first learned of PCB contamination and then, "if all PCBs had been cleaned up from New Bedford Harbor as of January first of this year," how often they would visit three particular beaches (Fort Phoenix Beach, East Beach and West Beach) in 1986. The final segments of the survey solicited standard demographic data.

3.2 Survey 2

Survey 2 was developed in May 1987 as a means of testing the reliability and validity of Survey 1. From the sample of 545 original respondents, 403 were successfully recontacted by MPR. Respondents were asked a question regarding recall of the previous survey. They were then asked if they spent any time at Fort Phoenix Beach, East Beach or West Beach in 1986. If they answered affirmatively, they were asked how often they had visited each of the three beaches between April and December, 1986.

Based on their phone number (ending in an odd or even digit), the respondents were split into two groups. Individuals in the first group were asked if they planned to visit the three named beaches in 1987 and, if yes, how many times they planned to visit each. Members of the second group were asked how often they had visited each of the three beaches since January first of the current year, how often they planned to visit each of the three beaches between the time of the interview (May) and Labor Day, and how often they planned to visit each of the three beaches between Labor Day and the end of the calendar year 1987.

All participating respondents were asked whether they felt that the water at the three beaches was safe for swimming. Those who responded affirmatively were only asked a series of concluding demographic questions. Those who felt that the water was unsafe for swimming were asked to elaborate an open-ended question: "What do you feel makes the water unsafe?" If PCBs were not listed among the reasons for the water being considered unsafe, respondents were queried as to whether the unsafe water was perhaps caused by PCBs or by something else. Respondents who still did not mention PCBs were again only asked the concluding demographic questions.

All respondents who by this point in the survey had mentioned PCBs (either voluntarily or in response to the direct question) were asked, "Suppose all the PCBs had been removed at the beginning of last year. Would you have gone to these beaches more often in 1986?" If the answer was positive, individuals were asked how many more times they would have visited each of the three beaches in question. Another hypothetical question was then presented regarding increased beach attendance in the event of PCB removal, but with other contaminants remaining in the water. The survey concluded with the standard demographic questions.

3.3 Survey 3

A second survey was conducted by MPR in September of 1987 to test the accuracy of respondents' predictions regarding beach use in the summer of 1987. In Survey 3, 342 interviews were completed and two more were partially completed.

Survey 3 began with a number of questions on actual beach attendance at East, West and Fort Phoenix Beaches during the summer of 1987 (between Memorial Day and Labor Day) and feelings regarding safety of the water for swimming. The survey again concluded with the standard demographic questions.

4 Errors and Biases

The CUM is subject to many of the same errors and biases to which the CVM can fall prey. These include errors and biases associated both with surveys generally and with hypothetical questions specifically. Indeed, many of the leading guidebooks and lists of "how-to" rules for designing questionnaires admonish against hypothetical questions altogether (Dillman 1978, 96). Hypothetical questions (such as those regarding future beach attendance in the event of cleanup) may produce careless responses, called "nonattitudes" by Converse (1974), involve little mental effort on the part of respondents (Feenberg and Mills 1980; Freeman 1979), or may result in responses designed solely to please the interviewer (DeLamater 1982; Mitchell and Carson, 1989). Where there is no cost for being wrong, there may be no incentive to be accurate.[16]

Accuracy can be a problem with respect to recalling events or activities as well as predicting events or activities, even under non-hypothetical conditions. The survey literature suggests that in both cases there is an upward bias for desirable activities and a downward bias for undesirable activities (Mitchell and Carson 1988). Chase and Harada (1983), in particular, find that individuals generally overestimate their participation in recreational activities. We consider first divergences between predicted and recalled beach usage in the New Bedford surveys.

4.1 Predicted Versus Recalled Usage

The 495 respondents to Survey 1 used in the government's analysis were asked how many times they had visited various New Bedford area beaches in 1985 and how many times they planned to visit those beaches in 1986.[17] All 363 respondents to Survey 2 who were also used in the government's analysis were asked whether they had spent any time in 1986 at East Beach, West Beach or Fort Phoenix State Beach.[18] If they answered yes, they were then asked for the number of times that they had visited each beach in 1986.[19] Since the government aggregated predictions for East and West beaches Table 1 shows predicted and recalled demand for East and

[16] Harris, Driver and McLaughlin (1989), drawing from recent psychology literature on human decision-making, isolate as one criterion for sound and reasonable decision-making the level of effort needed to determine a range of feasible alternatives. Other criteria include whether the objectives were clearly specified, and the degree of weighing of known costs and benefits of the alternatives.

[17] Originally Survey 1 began with a sample size of 545 respondents. Of these, 7 failed to answer the relevant part of the questionnaire and 43 failed to provide usable responses for beach attendance, bringing the working sample down to 495. Surveys 2 and 3 had a total sample of 403 and 342, respectively. Some of these were in the group of 50 original respondents not included in the government's working sample.

[18] Of the 403 respondents to Survey 2, 40 were not included amongst the 495 actually used in the government's analysis. Thus, the working sample for Survey 2 is 363.

[19] Predicted and recalled usage for individual beaches may not sum to 495 and 363, respectively, because of missing data. In aggregating individual beach usage, however, missing data were treated as zero values as long as a value was reported for at least one beach. All references in the text are to aggregate beach use unless otherwise noted.

West beaches combined, for Fort Phoenix State Beach and for all three beaches combined.

TABLE 1: Predicted and Recalled Usage, 1986								
	predicted 1986 usage (Survey 1, 1986, n=495)				recalled 1987 usage (Survey 2, 1987, n=363)			
	respondents predicting zero usage	respondents predicting positive usage	mean unconditional prediction	mean positive prediction	respondents recalling zero usage	respondents recalling positive usage	mean unconditional recall	mean positive recall
East/West	360 (73%)	135 (27%)	5.76	21.11	295 (81%)	68 (19%)	2.79	14.90
Ft. Phoenix	376 (76%)	119 (24%)	1.71	7.14	310 (86%)	51 (14%)	1.57	11.14
All Beaches	312 (63%)	183 (37%)	7.47	20.22	277 (76%)	86 (24%)	4.38	18.39

The appropriate statistical test for comparing the predicted mean to the recalled mean for individual beaches or for all beaches combined takes into account that the samples were not independent; that is, everyone who was in Survey 2 was also in Survey 1. Using the appropriate paired *t*-test on the 363 individuals who were in both surveys, we reject equality of the overall predicted and recalled means at the 99 percent significance level. However, of the 495 individuals who were in Survey 1, 312 (63 percent), predicted no beach usage in 1986 and of the 363 individuals who were in Survey 2, 277 (76 percent), recalled no beach usage in 1986. Conditional on positive usage, mean predicted usage for 1986 was 20.22 and mean recalled usage was 18.39. Using the appropriate paired *t*-test, we cannot reject equality of these means at any conventional significance level. We therefore isolate respondents who either predicted or recalled nonzero beach usage in 1986, denoting them as "users in 1986." As shown in Table 2, of the 363 individuals who were in both surveys, 204 were non-users in 1986.

TABLE 2: Predicted Versus Recalled Usage, 1986: Users and Non-Users (n=363)		
	respondents predicting positive usage	respondents predicting zero usage
respondents recalling positive usage	59 (16.3%)	27 (7.4%)
respondents recalling zero usage	73 (20.1%)	204 (56.2%)

As Table 2 also shows, of the 159 users in 1986, 73 predicted positive usage in 1986 but recalled zero usage in 1986, 59 predicted and recalled positive usage in 1986, and 27 predicted zero usage in 1986 but recalled positive usage in 1986. Thus, 72.5 percent of the respondents correctly predicted whether they would be users in 1986.

As Table 3 shows, the remaining respondents, roughly three-quarters overpredicted usage in 1986 and one-quarter underpredicted usage in 1986. Indeed, of all users, 109 overpredicted usage, two predicted and recalled the same positive usage, and 48 underpredicted usage.

TABLE 3: Underprediction Versus Overprediction, 1986 (n=363)

	correct prediction: zero usage	correct prediction: positive usage	underpredicted	overpredicted
East/West	240 (66%)	5 (1%)	40 (11%)	78 (21%)
Ft. Phoenix	259 (72%)	4 (1%)	32 (9%)	66 (18%)
All Beaches	204 (56.2%)	2 (.6%)	48 (13.2%)	109 (30%)

As mentioned above, Survey 2 respondents were divided into two groups (according to whether their phone number ended in an odd or an even digit). There were 179 and 177 individuals in each group, respectively. Respondents with even phone numbers were asked to recall beach usage for January through May of 1987 and to predict beach usage for June through September of 1987 and October through December of 1987 separately. Respondents with odd phone numbers were asked simply to estimate beach usage for all of 1987. In Survey 3, respondents with even phone numbers who were recontacted were asked to recall beach usage from June through September of 1987. This design allowed for a test of predicted versus recalled usage where the data on recalled usage were collected immediately after the period of usage, not many months later, as in Survey 2.

Tables 4, 5 and 6 replicate Tables 1, 2 and 3 for those individuals who had even phone numbers on Surveys 2 and 3. Any respondent who either predicted or recalled nonzero beach usage in June through September of 1987 is denoted as a user in 1987. In the 1987 data (Surveys 2 and 3) we again reject equality of the overall predicted and recalled means, but conditional on positive usage cannot do so. This pattern of results is identical to that found using the 1986 data.[20]

TABLE 4: Predicted and Recalled Usage, May-September, 1987[21]

	predicted 1987 usage (Survey 2, 1986, n=179)				recalled 1987 usage (Survey 2, 1987, n=146)			
	respondents predicting zero usage	respondents predicting positive usage	mean unconditional prediction	mean positive prediction	respondents recalling zero usage	respondents recalling positive usage	mean unconditional recall	mean positive recall
East Beach	135 (77%	40 (23%)	1.53	6.71	137 (94%)	9 (6%)	0.67	8.22
West Beach	142 (52%)	32 (18%)	1.15	6.23	139 (95%)	7 (5%)	0.67	12.18
Ft. Phoenix	130 (75%)	44 (25%)	1.78	7.05	133 (91%)	13 (9%)	0.66	7.81
All Beaches	117 (66%)	60 (34%)	4.40	12.97	124 (85%)	22 (15%)	1.75	11.64

Of the 146 individuals who were in Survey 2 and had even phone numbers in Survey 3, 89 were non-users in 1987. Of the 57 users in 1987, 35 predicted positive usage in 1987 but recalled zero usage in 1987, 17 predicted and recalled positive usage in 1987 and 5 predicted usage in 1987 but recalled positive usage in 1987. Again, this pattern of results is almost identical (in percentage terms) to 1986.

[20] Besides finding that individuals generally overestimate their participation in recreational activities, a result which we have confirmed, Chase and Harada (1984) also found that self-reports of recreational activity become particularly unreliable if the self-reports are removed from the reporting period by more than 60 days. Since Survey 2 was removed from the end of calendar year 1986 by over four months, and from the end of the primary beach use season in 1986 by approximately eight months, one may justifiably question the reliability of recalled beach usage for 1986 obtained by it. One purpose of Survey 3 was to compare predicted and recalled usage when there was essentially no lag between the end of the reporting period and self-reports of beach usage. Since Survey 3 also concentrated on the primary beach use season instead of the whole year, to test a "split ballot sample" technique was used to guarantee that the results on 1987 would be comparable to the results for 1986. Respondents to Survey 2 with even phone numbers were asked to estimate 1987 beach usage in three segments, January through April, May through September, and October through December. Respondents to Survey 2 with odd phone numbers were asked to estimate 1987 beach usage in only two segments, January through April, and May through December. There were no statistically significant differences between the overall beach usage in 1987 of respondents with even phone numbers and respondents with odd phone numbers. Neither were there any statistically significant differences between the distribution of users and nonusers, the mean of overall unconditional beach usage, or the mean of overall beach usage conditional on positive usage for 1987 as compared to 1986.

[21] The total sample size for respondents recalling beach use in 1987 was 146. This constitutes the subset of the 177 respondents to Survey 2 with odd phone numbers who were recontacted in Survey 3. We again note that individual predicted and recalled beach usage may total to less than 179 and 146, respectively, due to missing data.

TABLE 5: Predicted Versus Recalled Usage, May-September 1987: Users and Non-Users (n=146)		
	respondents predicting positive usage	respondents predicting zero usage
respondents recalling positive usage	17 (11.7%)	5 (3.4%)
respondents recalling zero usage	35 (24%)	89 (61%)

Of all users in 1987, 44 overpredicted usage, 3 predicted and recalled the same positive usage and 10 underpredicted usage, a pattern once again almost identical (in percentage terms) to 1986.

TABLE 6: Underprediction Versus Overprediction, 1987 (n=146)				
	correct prediction: zero usage	correct prediction: positive usage	underpredicted	overpredicted
East Beach	105 (72%)	0	8 (6%)	32 (22%)
West Beach	111 (77%)	0	5 (3%)	28 (19%)
Ft. Phoenix	103 (72%)	2 (1%)	5 (3%)	34 (24%)
All Beaches	89 (61%)	3 (2%)	10 (7%)	44 (30%)

The data for 1987 confirms the result obtained from the 1986 data that individuals are relatively accurate in predicting whether they will use any beach at all. We next ask whether individuals who either predicted or recalled positive beach usage in 1986 -- *i.e.*, were users in 1986 -- also predicted or recalled positive beach usage in the summer of 1987 -- *i.e.*, were users in 1987. Table 7 answers this question.

TABLE 7: Users and Non-Users in 1986 Versus Users and Non-Users in 1987 (n=146)		
	user in 1987	non-user in 1987
user in 1986	50 (34%)	21 (14%)
non-user in 1986	7 (5%)	68 (47%)

It is clear from Table 7 that respondents were relatively consistent in their use/non-use decisions -- only 28 of 146 respondents for whom we are able to compare 1986 and 1987 switched between use and non-use. However, among respondents who were users in both years, the pattern of underpredictions and overpredictions shows no temporal consistency. Table 8 shows the relationship between underpredicting, predicting correctly, and overpredicting for respondents who predicted or recalled positive beach usage in both 1986 and 1987.

TABLE 8: Underprediction and Overprediction in 1986 Versus Underprediction and Overprediction in 1987 (n=50)			
	underprediction in 1987	correct prediction in 1987	overprediction in 1987
underprediction in 1986	2	0	14
correct prediction in 1986	0	0	1
overprediction in 1986	6	3	24

Together Tables 7 and 8 reveal a striking pattern: respondents are very good at predicting any usage (measured relative to recalling any usage), but for those who do predict or recall positive usage, the pattern of overpredicting versus underpredicting is basically random. This means that while one might expect to be able to correlate the decision to use beaches at all with other survey responses (such as PCB awareness or responses to hypothetical changes in PCB levels), it is unlikely that the pattern of underprediction versus overprediction for users will be systematically related to anything else.

The implications of these results for the CUM are significant and disturbing. While individuals in the New Bedford survey were relatively accurate with respect to predicting any actual beach usage, they tended to overpredict actual beach usage conditional on predicted or recalling some beach usage. However, the pattern of overprediction versus underprediction was not stable over time. Thus, it is difficult to envision accurate responses with respect to hypothetical beach usage.

4.2 PCB Awareness Measures

As a prelude to estimating the effects of removing PCB's from the New Bedford Harbor on beach usage, Survey 1 attempted to identify those individuals who were "aware" of PCBs. One of the major purposes of Survey 2 was to show that the procedure used to elicit PCB awareness in Survey 1 was biased in two fundamental ways. First, the question preceding the questions used to elicit PCB awareness may have inappropriately influenced respondents; and, second, the questions used to elicit PCB awareness were themselves lacking in objectivity.

The first question in Survey 1 related to environmental quality of the New Bedford Harbor solicited a subjective rating:

> On a scale from 1 to 10 where '10' is 'excellent,' and '1' is 'very poor,' how would you rate the environmental quality of New Bedford Harbor?

While self-rating scales are simple to operate and may be more informative in some contexts than straight yes/no answers, they are entirely subjective and may be of little use for complex and emotion-laden subjects (Moser 1958, 236). Indeed, leading survey methodologists such as Dillman (1978) have questioned the acceptability of scaling techniques for obtaining graduated responses, particularly in surveys of the general public. More significantly, the scaling question in Survey 1 regarding environmental quality was worded so as to present the numerical equivalent of the very poor rank just before the query on environmental quality ("...'very poor,' how would you rate the environmental quality...?"). It has been demonstrated that, due to the recency effect, in verbal (as opposed to written) presentation of a scale, the last mentioned alternative tends to be favored (Kalton, Collins and Brook 1978, 155).[22]

The sequencing of questions and lead-in to the substance of Survey 1 may have biased responses on overall environmental quality toward the poor end of the scale and may have "primed" the responses to the follow-on questions on harmful substances and PCBs. Indeed, the next question on Survey 1 asked the respondent "what specific substances or chemicals, if any, do you think are damaging the environmental quality of the harbor?" Responses included PCBs, other chemicals or substances besides PCBs, none, or "don't know." Respondents who mentioned other chemicals or substances besides PCBs or answered "don't know" were then asked "do you believe the harbor is contaminated with PCBs?"

Following the pattern of the questions, we define four "PCB awareness" categories for respondents to Survey 1: (1) "no PCB awareness" -- no stated awareness of PCBs; (2) "strong PCB awareness" -- referred to PCB damage without prompting; (3) "moderate PCB awareness" -- referred to damage from substances or chemicals besides PCBs and then answered yes to the prompt regarding PCB contamination; and (4) "weak PCB awareness" -- answered "don't know" to the open question on substances or chemicals

---

22 Rotation of answer choices is a relatively simple way of minimizing this bias (Dillman 1978, 216).

damaging the harbor and then answered yes to the prompt regarding PCB contamination. Table 9 shows the distribution of respondents to Survey 1 according to these categories of PCB awareness.

TABLE 9: PCB Awareness Distribution: Survey 1, 1985 (n=495)			
not PCB aware	strong PCB awareness (no prompt)	moderate PCB awareness ("other" prompt)	weak PCB awareness ("don't know" prompt)
109 (22%)	206 (41.7%)	100 (20.2%)	80 (16.2%)

In Survey 1, only 22 percent of the respondents failed to agree in one way or another that PCBs were damaging New Bedford Harbor. But those respondents who agreed that PCBs were damaging New Bedford Harbor may have been biased by the form of the questions on harmful substances and PCBs. For example, those respondents who agreed that PCBs were present when prompted may have been susceptible to acquiescence (DeLamater 1982, 27) or "yea-saying;" the propensity to agree with an interviewer's request regardless of one's true views (Mitchell, Carson 1988, 240).[23]

A "don't know" response may also be a way of refusing to give a definite answer, indicate genuine lack of knowledge, a fear of giving the "wrong" answer, an inability to decide, a failure to understand the question, or a lack of interest (Moser 1958, 189). Or, it may suggest an unwillingness to do the mental work required to give an answer (Fowler 1984, 81). The follow-up questions to the open-ended question regarding chemicals or other substances that may be damaging the environmental quality of New Bedford Harbor may have challenged some respondents to expend the mental effort to give a response, but at the same time they may have inherently biased the response toward a specific answer.[24]

---

23 Moser (1958, 225) identifies the problem of a leading question which by "its content, structure or wording, leads the respondent in the direction of a certain answer" or which "suggests only some of the possible answers."

24 Closed-end questions such as this one where the alternatives are incomplete offer "answer-suggestions," and amount to "nearly an answer-coercion" (Molenaar 1982, 56). Although closed-ended questions are useful in guiding the respondent's search for an answer, dichotomous choice or yes/no questions with only one predicate or alternative are imbalanced in the sense that an alternative tends to be chosen more when it is presented alone rather than with other options (Molenaar 1982, 58) and with a strong bias in the direction of the affirmative response (Kalton, Collins, Brook 1978, 150).

Survey 2 was designed in part to demonstrate the problematic nature of the PCB awareness categories defined in Survey 1. The first question in Survey 2 concerned with environmental quality of the New Bedford Harbor asked respondents "Do you feel the water at East, West and Fort Phoenix Beaches is safe to swim in?" If a respondent answered "yes," the respondent subsequently was asked only the standard set of demographic questions. If the respondent answered "don't know," the respondent was next asked an open-ended question: "What makes you unsure of whether the water is safe to swim in?" If PCBs were not mentioned, the respondent subsequently was asked only the standard set of demographic questions. If the respondent answered "no" regarding water safety for swimming, the respondent also was asked an open-ended question: "What do you feel makes the water unsafe?" In this case, if the respondent failed to mention PCBs, a prompt was employed asking the respondent whether the reason they felt the water at New Bedford beaches is unsafe for swimming was caused by PCBs as opposed to something else. If a respondent answered "something else" they subsequently were asked only the standard set of demographic questions.

Again following the pattern of the questions, we define four "PCB awareness" categories for respondents to Survey 2: (1) "no PCB awareness" -- no stated awareness of PCBs; (2) "strong PCB awareness" -- felt the water was unsafe for swimming and mentioned PCBs as a cause without a prompt; (3) "moderate PCB awareness" -- felt the water was unsafe for swimming and did not mention PCBs directly as a cause, but responded affirmatively to the prompt regarding PCB contamination; and (4) "weak PCB awareness -- answered "don't know" to the question regarding the safety of water at New Bedford Beaches for swimming and then mentioned PCBs in response to the question "why." Table 10 shows the distribution of respondents to Survey 2 according to these categories of PCB awareness.

TABLE 10: PCB Awareness Distribution Survey 2, 1986 (n=363)

not PCB aware	strong PCB awareness ("unsafe" no prompt)	moderate PCB awareness ("unsafe" prompt)	weak PCB awareness ("don't know")
201 (55%)	92 (25%)	55 (15%)	15 (4%)

Survey 2 generated a much lower yield of PCB-aware respondents than Survey 1 -- 55 percent failed to agree in one way or another that PCBs were making water at the New Bedford Beaches unsafe for swimming. In Survey 1, only 22 percent of all respondents failed to agree in one way or another that PCBs were damaging New Bedford Harbor.

Table 11 provides a comparison of PCB awareness measures across the two surveys for those respondents who answered both surveys. Of the 363 respondents who were in both surveys, the majority who did not state any PCB awareness in Survey 1 (74 of 85) also revealed no PCB awareness in Survey 2. However, an additional 127 respondents who had suggested some form of PCB awareness in Survey 1 did not state any PCB awareness in Survey 2. In fact, 51 of the 201 respondents who indicated no PCB awareness in Survey 2 indicated an awareness of PCB damage without a prompt in Survey 1. Only 67 of the 152 respondents who had an awareness of PCB damage without a prompt in Survey 1 again mentioned PCBs as a reason why the New Bedford Beaches were unsafe for swimming without a prompt in Survey 2.

TABLE 11: Comparison of PCB Awareness Measures -- Survey 1 Versus Survey 2 (n=363)

	not PCB aware	strong PCB awareness ("unsafe" no prompt)	moderate PCB awareness ("unsafe" prompt)	weak PCB awareness ("don't know")	row total
not PCB aware	74 (87%)	5 (6%)	6 (7%)	0	85
strong PCB awareness (no prompt)	51 (34%)	67 (44%)	24 (16%)	10 (7%)	152
moderate PCB awareness ("other" prompt)	41 (55%)	14 (19%)	16 (22%)	3 (4%)	74
weak PCB awareness ("don't know" prompt)	35 (67%)	6 (12%)	9 (17%)	2 (4%)	52
column total	201	92	55	15	363

While Table 11 makes it apparent that Survey 1 and Survey 2 elicited very different types of PCB awareness, the question remains uncertain whether there was any systematic relationship between beach use and PCB awareness levels. Tables 12 and 13 show the relationships between users in 1986 and the four PCB awareness categories for Survey 1 and users in 1987 and the four PCB awareness categories for Survey 2, respectively.

TABLE 12: Survey 1 PCB Awareness Distribution by Users and Non-Users in 1986 (n=363)	not PCB aware	strong PCB awareness (no prompt)	moderate PCB awareness ("other" prompt)	weak PCB awareness ("don't know" prompt)	row total
non-user in 1986	56 (27%)	79 (39%)	35 (17%)	34 (17%)	204
user in 1986	29 (18%)	73 (46%)	39 (25%)	18 (11%)	159
x^2 (1 degree of freedom)	4.23*	1.90	2.99**	2.08	363

* Significant at the five-percent level
** Significant at the ten-percent level

TABLE 13: Survey 2 PCB Awareness Distribution by Users and Non-Users in 1987 (n=146)	not PCB aware	strong PCB awareness ("unsafe" no prompt)	moderate PCB awareness ("unsafe" prompt)	weak PCB awareness ("don't know")	row total
non-user in 1987	56 (63%)	17 (19%)	11 (12%)	5 (6%)	89
user in 1987	29 (51%)	18 (32%)	9 (16%)	1 (2%)	57
x^2 (1 degree of freedom)	2.07	2.97*	.35	1.32	146

* Significant at the one-percent level

While users in 1986 were significantly more likely than non-users to show some PCB awareness, users in 1986 were not significantly more likely than non-users to show the strong form of PCB awareness in Survey 1. The opposite pattern of effects holds in 1987 -- while overall users in 1987 were not significantly more likely than non-users to show some form of PCB awareness, users in 1987 were significantly more likely to show the strong form of PCB awareness. Also, with respect to the moderate form and the weak form of PCB awareness in Survey 1 and Survey 2, there is a statistically significant tendency for users in 1986 to favor the moderate form of PCB awareness in Survey 1.

Thus, the order of questions and the procedure used to elicit PCB awareness in Survey 1 lead respondents who predicted or recalled positive usage in 1986 to be more likely to state some form of PCB awareness than those who both predicted and recalled zero usage in 1986. Yet the only form of PCB awareness in Survey 1 sensitive to the distinction between users and non-users was one in which the respondent initially referred to damage from substances or chemicals besides PCBs and only mentioned PCBs in response to a prompt. The order of questions and procedure used to elicit PCB awareness in Survey 2 lead respondents who predicted or recalled positive usage in 1987 to be no more likely than those who both predicted and recalled zero usage in 1987 to indicate some form of PCB awareness. However, in this latter case, the only form of PCB awareness that was sensitive to the distinction between users and non-users was one in which the respondent felt the water was unsafe for swimming and mentioned PCBs as a cause without a prompt.

These results strongly suggest that procedures which elicit statistically significant unbiased responses to questions regarding awareness of environmental damages are, under the best circumstances, difficult to devise. However, the analysis also suggests that the format of the question is extremely important.

4.3 Hypothetical Demand

One of the major purposes of the government survey was to estimate the lost use value to beach users of New Bedford beaches stemming from PCB contamination. To accomplish this the following question was asked of all respondents to Survey 1 who indicated some form of PCB awareness: "If all PCBs had been cleaned up from New Bedford Harbor as of January 1st of this year, how often would you visit the following beaches [East Beach, West Beach and Fort Phoenix State Beach] during 1986?" By comparing the answer to this question to predicted demand, those respondents who would increase their usage of each of the beaches under the hypothetical "cleanup" scenario are identified. We also aggregate across the three beaches to identify those respondents who would have increased their overall beach usage had all three beaches been cleaned up. With respect to the latter, of the 386 respondents to Survey 1 who indicated some form of PCB awareness, 200 indicated that they would increase their usage of New Bedford Beaches if PCBs were removed.

There are a number of problems with the manner in which Survey 1 attempted to elicit information regarding respondents' hypothetical beach usage in the event of no PCB contamination. The question regarding hypothetical beach usage was asked only of respondents who indicated PCB awareness. However, as we have already discussed, the manner in which Survey 1 categorized respondents as PCB aware had a number of problems.[25] Perhaps even more significantly, the question about hypothetical beach usage is itself ill-posed (Cross 1989, 319). In particular, no detailed scenario is developed as a context in which the respondent can place the proposed hypothetical commodity, in this case a PCB-free beach. The term "cleanup" is not defined, leaving each respondent to make the term meaningful for himself or herself given the respondent's own frame of reference. The lack of visual cues forces respondents to determine their own geographic boundaries for the hypothetical commodity. No baseline level of existing PCB contamination is established, nor are meaningful increments of change used to define the hypothetical commodity.[26]

Survey 2 approached the issue of hypothetical demand at the New Bedford beaches in the absence of PCB contamination in a manner only slightly different from Survey 1. Respondents who indicated some PCB awareness in Survey 2 were asked the following question: "Suppose all the PCBs had been removed at the beginning of last year. Would you have gone to these beaches [East Beach, West Beach and Fort Phoenix] more often in 1986?" Respondents who answered yes were then asked specifically how many more times they would have visited East Beach, West Beach and Fort Phoenix Beach. Of the 162 respondents who indicated some PCB awareness in Survey 2, 82 indicated that they would not have gone to these beaches more often had PCBs been removed, 72 indicated that they would have gone more often, and six answered, "don't know."

In particular, we estimated four conditional logit models in which the independent variable is either the increase in use of PCB-free beaches in Survey 1 or the increase in use of PCB-free beaches in Survey 2, and the dependent variables are the various PCB awareness categories.[27] The results of these logit models are shown in Tables 14 and 15. The within-survey results are extremely weak. There is no statistical

---

[25] Specifically, it tends to be biased toward the population of beach users. Given that the results of the survey were used to estimate demand curves for beach usage, this amounts to sampling based on the dependent variables, a well-known error in statistical analysis.

[26] An example of the latter is provided by Desvousges, Smith and McGiveny (1983), who, in their study of the Monongahela River, asked respondents about changes in water quality ranging from boatable to swimmable to drinkable. Mitchell and Carson (1989, 184) go so far as to cite "cut down" in air pollution and "cleanup" of water pollution as instances of improperly vague descriptions of goods in survey instruments.

[27] The use of conditional logistic models in this context is equivalent to the estimation of standard log-linear models wherein some higher order interactions are set a *priori* to zero.

relationship between a respondent's category of PCB awareness in Survey 1 and the respondent's answer to the hypothetical question in Survey 1 about increased use of PCB-free beaches. The only statistically significant relationship between a respondent's category of PCB awareness in Survey 2 and the respondent's answer to the hypothetical question in Survey 2 about increased use of PCB-free beaches is a negative relationship between the weak form of PCB awareness and increased demand for PCB-free beaches.

TABLE 14: Conditional Logit Models of Increased Demand for Beaches Under Hypothetical No-PCBs Scenario: Effect of Survey 1 PCB Awareness Categories

independent variables	dependent variables	
	increase use (Survey 1)	increase use (Survey 2)
	n = 386	n = 164
not PCB aware in Survey 1	N.A.	-.405 (-.768)
strong PCB awareness (no prompt) in Survey 1	.030 (.280)	-1.372 (-1.81)
moderate PCB awareness ("other" prompt) in Survey 1	.241 (1.20)	.238 (.690)
weak PCB awareness ("don't know" prompt) in Survey 1	-.050 (-.225)	.357 (.725)

TABLE 15: Conditional Logit Models of Increased Demand for Beaches Under Hypothetical No-PCBs Scenario: Effect of Survey 2 PCB Awareness Categories

independent variables	dependent variables	
	increase use (Survey 1)	increase use (Survey 2)
	n = 278	n = 156
not PCB aware in Survey 2	.016 (.089)	N.A.
strong PCB awareness ("unsafe" no prompt) in Survey 2	.641 (2.84)	-.158 (-.741)
moderate PCB awareness ("unsafe" prompt) in Survey 2	.122 (.428)	.077 (.277)
weak PCB awareness ("don't know") in Survey 2	-1.39 (-2.15)	-1.01 (-1.73)

Only slightly stronger results are obtained in cross-survey comparisons. With respect to the relationship between a respondent's category of PCB awareness in Survey 1, if any, and increased use of PCB-free beaches in Survey 2, we find a negative relationship between the strong form of PCB awareness in Survey 1 and increased use of PCB-free beaches in Survey 2. With respect to the relationship between a respondent's category of PCB awareness in Survey 2 and the increased demand for PCB-free beaches in Survey 1, we find a positive relationship between the strong form of PCB awareness in Survey 2 and a negative relationship between the weak form of PCB awareness in Survey 2 and increased demand for PCB-free beaches in Survey 1. Overall then, Table 14 and 15 suggest that PCB-awareness categories are not closely linked to hypothetical changes in demand. Specifically, they provide additional evidence that the weak form of PCB awareness, related to "don't know" responses, is a particularly poor measure of PCB awareness.

4.4 Amenity Misspecification

As noted in Section 4.3, the elicitation procedure used in Survey 1 can lead to a form of bias called amenity misspecification. In this type of bias, the respondent values a perceived good which is different than the researcher's intended good. According to Mitchell and Carson (1990), "the description of the good in CVM surveys typically contains several elements such as the time period within which the good is to be provided; the location, cause and size of the change; and the nature of the amenity itself" (p. 249). That is, a contingent commodity is most appropriately defined in the context of an entire "scenario." A lack of familiarity with or a lack of understanding of the scenario in which a hypothetical commodity is characterized, particularly one that is ill-defined, may lead respondents to rely on judgmental heuristics, distorting the survey results. In other words, since people may not have previous well-defined values for a particular good, it is likely that they may ignore or distort some or all of the details of a scenario.[28]

Survey 2 attempted to address the possibility of amenity misspecification in Survey 1 by asking about planned beach attendance in the event of two scenarios: First, PCB removal from the Harbor and then PCB cleanup, but with all other contaminants remaining in the water. In fact, the first question in Survey 2 regarding hypothetical demand for PCB-free beaches deliberately was worded in a fashion similar

---

[28] According to Tversky and Kahneman (1974, 1124-1131), people rely on a limited number of heuristic principles to reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. The use of these heuristic principals inevitably introduces distortions of one form or another compared to the full-information, full-optimization outcome.

to the analogous question in Survey 1. However, in Survey 2 respondents who said that they would use PCB-free beaches more often were then asked:

> Now suppose all of the PCBs had been removed at the beginning of last year, but other contaminants in the water were not removed. Would you have gone to these beaches more often in 1986?

The results of this "test" for amenity misspecification are shown in Table 16. The ratio of respondents who "changed their mind" to those who continued to say that they would use PCB-free beaches more even if all other contaminants remained was approximately 4 to 1 (58 versus 15). Furthermore, of the 20 respondents who stated that they would not increase their use of PCB-free beaches in Survey 1 but that they would in Survey 2, 18 changed their minds when reminded that other contaminants would remain. The "test" is thus rather conclusive -- Survey 1 suffers badly from amenity misspecification.

TABLE 16: Relationship Between Increased Demand for Beaches Under Hypothetical No-PCBs Scenario and "Amenity Misspecification" (n=146)

	increase in use given PCBs only (Survey 2)	no increase in use given PCBs only (Survey 2)
increase in use given no-PCBs (Survey 1)	13	34
no increase in use given no-PCBs (Survey 1)	2	18
increase in use given no-PCBs (Survey 2)	15	58

Tables 17 and 18 report the results of two logit models which relate PCB awareness categories to the amenity misspecification question. Respondents who were PCB-aware according to Survey 1 were less likely to say they would use PCB-free beaches more, even if all other contaminants remained, with the order of significance of the effect strongest for the strong form of PCB awareness and weakest for the weak form of PCB awareness. A similar pattern of effects holds for the PCB awareness categories as defined in Survey 2, except that the weak form of PCB awareness is insignificant. In other words, respondents who were most "aware" of PCBs (by either Survey 1 or Survey 2 definition) were most likely to be "aware" of other contaminants as well and thus most likely to change their minds regarding increased use of PCB-free beaches when reminded that other contaminants would remain.

TABLE 17: Conditional Logit Models: "Amenity Misspecification" as a Function of Survey 1 PCB Awareness Categories: Dependent Variable = Increased Demand for Beaches Given PCBs Only Removed	
independent variables	coefficients (t-statistics)
(n=73)	
not PCB aware in Survey 1	-10.18 (-.153)
strong PCB awareness (no prompt) in Survey 1	-1.17 (-3.06)
moderate PCB awareness ("other" prompt) in Survey 1	-1.32 (-2.35)
weak PCB awareness ("don't know" prompt) in Survey 1	-1.38 (-1.75)

TABLE 18: Conditional Logit Models: "Amenity Misspecification" as a Function of Survey 2 PCB Awareness Categories: Dependent Variable = Increased Demand for Beaches Given PCBs Only Removed	
independent variables	coefficients (t-statistics)
(n=73)	
strong PCB awareness ("unsafe" no prompt) in Survey 2	-1.10 (-3.02)
moderate PCB awareness ("unsafe" prompt) in Survey 2	-1.48 (-2.99)
weak PCB awareness ("don't know") in Survey 2	-8.64 (-.20)

5 Summary of the Results and Policy Recommendations

With respect to the New Bedford Harbor CUM study, the following conclusions can be drawn from our analysis.

a. **Bias Toward Overprediction.** Respondents are relatively accurate in predicting whether or not they will use the New Bedford Harbor beaches. These predictions are also stable over time. However, among those who predict or recall nonzero beach usage, there is a bias in favor of overprediction. No relationship exists between individuals who overpredict in one season and those who overpredict in other seasons.

b. **Time-Span Effect.** No observable effects result from asking respondents to estimate beach usage for an entire year versus estimating beach usage for three subperiods of the year.

c. **Leading Questions.** By using leading questions and strong prompting, it is possible to elicit substantial "PCB awareness" among respondents. However, PCB awareness elicited by these techniques does not correlate appropriately with beach use or non-use, nor does it explain subsequent responses to hypothetical questions regarding increased use of "PCB-free" beaches. More neutral elicitation techniques yield PCB awareness measures which are correlated more appropriately with beach use or non-use. Such measures, however, still do not explain responses to hypothetical questions regarding increased use of PCB-free beaches.

Both PCB awareness measures elicited using leading questions and strong prompting and PCB measures elicited using more neutral elicitation techniques are negatively related to the likelihood that a respondent who stated that he or she would use PCB-free beaches more subsequently indicates that he or she would still use the beaches more under a PCB-only scenario. In other words, those who are most aware of PCBs by either measure appear to be most aware of other chemicals and substances that might be damaging to the quality of the New Bedford area beaches. Thus questions related to hypothetical PCB removal are likely to elicit the most biased results form this group of respondents.

d. **PCB Removal Only.** When respondents who indicate that they would use PCB-free beaches more are subsequently reminded that only PCBs hypothetically are being removed from the harbors, the vast majority stated that they would not, in fact, use the beaches more under the "PCBs-only" scenario.

These results strongly suggest that in the New Bedford case using demand curves based on hypothetical usage to estimate the value of cleanup is highly sensitive to the nature of the survey instrument. In the original government survey, there was a positive relationship between beach users and PCB awareness as defined in that survey. However, users tend to overpredict beach usage under the *status quo*. Furthermore, the stronger the form of PCB awareness, the more likely the respondent will state that he or she would not use a beach more often if only PCBs are removed.

In the absence of a standard set of guidelines for administering CVM and CUM surveys, a potential for serious problems exists. Several attempts at establishing such guidelines have been made. Based on a comparison of studies using the CVM and drawing on analysis of market-like behavior in experimental economics and the psychology literature,[29] Cummings, Brookshire and Schulze (1986) proposed what they call Reference Operating Conditions (ROC) which should be satisfied by studies using the CVM techniques:

1. Subjects-participants in the CVM must understand (be familiar with) the commodity to be valued.

2. Subjects must have had (or be allowed to obtain) prior valuation and choice experience with respect to consumption levels of the commodity.

3. There must be little uncertainty.

4. Willingness-to-pay, not willingness-to-accept, measures are elicited.[30]

Three additional conditions were proposed by Daniel Kahneman (Cummings, Brookshire and Schulze, 1986, 186-194):

5. The CVM should only be used for problems that have a "purchase structure."

6. The use of the CVM should be restricted to user values, rather than to ideological values.

7. Accurate description of payment mode is essential to the CVM.

Kahneman's ROC's were themselves subsequently qualified by Cumming, Brookshire and Schulze in the conclusion to their book (1986, 230-231), as were the third and fourth of the original ROC. In fact, Mitchell and Carson (1989, 93-94) ultimately argued that to the extent that the first and second of the original four ROCs are based on the "consumer goods market model," they are also "inappropriate," proposing that "political markets are a more appropriate analogue for CVM surveys that value public goods than are private markets."

Based on the problems identified in this study, as well as the inability of the CVM practitioners to agree on even a basic set of criteria which would be met by studies using the CVM, it is arguable that the contingent techniques fall far short of satisfying the foundational criteria needed for survey evidence to be admissible. Problems with the CVM are now well-known. Despite their seeming objectivity, studies

---

29 Harris, Drive, McLaughlin (1989); Cummings, Brookshire, Schulze (1986); and Fischhoff, Slovic and Lichtenstein (1980).

30 WTP and WTA often diverge substantially. For a discussion of the problem and the literature associated with it, see Hoffman and Spitzer (1990).

using the CUM are similarly unlikely to be admissible as evidence. Given the inherent problems with both, as documented by case studies like this and that done by Randall, Hoehn and Tolley (1981), CVM and CUM studies may encounter admissibility problems in the courtroom and may also be judged unpersuasive by either a jury of objective economists or laymen.

Basing public policy decisions or damage awards on contingent usage studies can be as problematic as basing them on contingent valuation studies. While the latter are becoming increasingly criticized by much of the economics profession, the former have not yet received similar scrutiny. Contingent usage studies avoid some of the problems of contingent valuation studies, but ultimately they may be just as problematic with respect to their admissability as evidence.

REFERENCES

Bishop, Richard C., and Thomas A. Heberlein. "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?" American Journal of Agricultural Economics 61 5 (December 1979): 926-930.

Chase, David R., and Geoffrey C. Godbey. "The Accuracy of Self-Reported Participation Rates." Leisure Studies 2 (1983): 231-235.

Chase, David R., and Munehiko Harada. "Response Error in Self-Reported Recreation Participation." Journal of Leisure Research 16 (1984): 322-329.

Cicchetti, Charles J., and Neil Peck. "Assessing Natural Resources Damages: The Case Against Contingent Value Survey Methods." Natural Resources and Environment 4 1 (Spring 1989).

Cicchetti, Charles J., and V. Kerry Smith. The Costs of Congestion: An Econometric Analysis of Wilderness Recreation. Cambridge: Ballinger Publishing Co., 1976.

Converse, Phillip E. "Comment: The Status of Non-Attitudes." The American Political Science Review 68 2 (June 1974): 650-660.

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

DeLamater, John. "Response-Effects of Question Content." In Response Behaviour in the Survey-Interview, edited by Wil Dijkstra and Johannes van der Zouwen. London/New York: Academic Press, 1982.

Desvousges, William H., V. Kerry Smith, and Matthew P. McGiveny. "A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvements." EPA Contract Report 68-01-5838, prepared for Economic Analysis Division of United States Environmental Protection Agency, 1983.

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

Dillman, Don A. Mail and Telephone Surveys: The Total Design Method. New York: Wiley, 1978.

Feenberg, Daniel, and Edwin S. Mills. Measuring the Benefits of Water Pollution Abatement. New York: Academic Press, 1980.

Fischhoff, Baruch, Paul Slovic, and Sarah Lichtenstein. "Knowing What You Want: Measuring Labile Values." In Cognitive Processes in Choice and Decision Behavior, edited by Thomas S. Wallsten. Hillsdale, NJ: Lawrence Erlbarum Associates, 1980.

Fowler, Floyd J. Survey Research Methods. Beverly Hills, CA: Sage Publications, 1984.

Freeman, A. Myrick, III. "The Benefits of Air and Water Pollution Control: A Review and Synthesis of Recent Estimates." Report to the Council on Environmental Quality, Washington, DC, 1979.

Freeman, A. Myrick, III. The Benefits of Environmental Improvement Theory and Practice. Baltimore: The Johns Hopkins University Press for Resources for the Future, 1979.

Freeman, A. Myrick, III. "Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues." Scandinavian Journal of Economics 81 (1979): 154-173.

Granbois, Donald H., and John O. Summers. "Primary and Secondary Validity of Consumer Purchase Possibilities." Journal of Consumer Research 1 (March 1975): 31-38.

Green, Eric D., and Charles R. Nesson. Federal Rules of Evidence. Boston: Little Brown, 1988.

Harris, Charles C., B. L. Driver, and William J. McLaughlin. "Improving the Contingent Valuation Method: A Psychological Perspective." Journal of Environmental Economics and Management 17 3 (1989): 213-229.

Hoffman, Elizabeth, and Matthew L. Spitzer. "The Divergence Between Willingness-to-Pay and Willingness-to-Accept Measures of Value." Social Science Working Paper 755, California Institute of Technology, November 1990.

Huppert, Daniel D. "Measuring the Value of Fish to Anglers: Application to Central California Anadromous Species." Marine Resource Economics 6 (1989).

Juster, F. Thomas. "Consumer Anticipations and Purchase Probability: An Experiment in Survey Design." American Statistical Association Journal 61 (1966): 658-696.

Kain, John F., and John M. Quigley. "Note on an Owner's Estimate of Housing Value." Journal of the American Statistical Association 67 (December 1972).

Kalton, Graham, Martin Collins, and Lindsay Brook. "Experiments in Wording Opinion Questions." Applied Statistics 27 (1978): 155.

Kealy, Mary Jo, John F. Dovidio, and Mark L. Rockel. "Accuracy in Valuation is a Matter of Degree." Land Economics 64 2 (May 1988): 158-171.

Kealy, Mary Jo, Mark Montgomery, and John F. Dovidio. "Reliability and Predictive Validity of Contingent Values: Does the Nature of the Good Matter?" Journal of Environmental Economics and Management 19 (1990): 244-263.

Kish, Leslie, and John B. Lansing. "Response Error in Estimating the Value of Homes." Journal of the American Statistical Association 49 (September 1954): 520-538.

Kling, Catherine L. "The Reliability of Estimates of Environmental Benefits from Recreation Demand Models." American Journal of Agricultural Economics 70 (1988): 892-901.

Kopp, Raymond J., Paul R. Portney, and V. Kerry Smith. "An Economic Appraisal of the DC Appeals Court Ruling on the DOI Regulations for Natural Resource Damage Assessments." Washington, DC: Resources for the Future, 1990.

Loomis, John B. "Test-Retest Reliability of the Contingent Valuation Method: A Comparison of General Population and Visitor Responses." American Journal of Agricultural Economics 71 1 (February 1989): 76-84.

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

Miller, Peter V., and Robert M. Groves. "Matching Survey Reports to Official Records: An Exploration of Validity in Victimization Reporting." Public Opinion Quarterly 49 (1985): 366-380.

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

Molenaar, Nico, J. "Response-Effects of 'Formal' Characteristics of Questions." In Response Behaviour in the Survey-Interview, edited by W. Dijkstra and J. van der Zouwen, 49-89. London/New York: Academic Press, 1982.

Morgan, Fred W. "Judicial Standards for Survey Research: An Update and Guidelines." Journal of Marketing 54 (January 1990): 59-70.

Morrison, Donald G. "Purchase Intention and Purchase Behavior." Journal of Marketing 43 (Spring 1979): 65-74.

Moser, C. A. Survey Methods In Social Investigation. London: Heinemann Educational Books, Ltd., 1958.

Newman, Joseph W., and Bradley, D. Lockeman. "Measuring Prepurchase Information Seeking." Journal of Consumer Research 2 (December 1975): 216-222.

Parfitt, J. "A Comparison of Purchase Recall with Diary Panel Records." Journal of Advertising Research 8 (1968): 1631.

Parry, Hugh J., and Helen M. Crossley. "Validity of Responses to Survey Questions." Public Opinion Quarterly 49 (September 1954).

Randall, Alan, John P. Hoehn, and George S. Tolley. "The Structure of Contingent Markets: Some Experimental Results." Paper presented at the Annual Meeting of the American Economic Association, Washington, DC, December 1981.

Randall, Alan, and J. R. Stoll. "Existence Value in a Total Valuation Framework." In Managing Air Quality and Scenic Resources at National Parks and Wilderness Areas, edited by R. D. and L. G. Rowe. Boulder, CO: Westview Press, 1983.

Reiling, Stephen D., Kevin J. Boyle, Marcia L. Phillips, and Mark W. Anderson. "Temporal Reliability of Contingent Values." Land Economics 66 2 (May 1990): 128-134.

Schulze, William D., David S. Brookshire, and Mark A. Thayer. "National Parks and Beauty: A Test of Existence Values." Paper presented at the American Economic Association Annual Meeting, Washington, DC, 1981.

Smith, V. Kerry, William H. Desvousges, and Ann Fisher. "A Comparison of Direct and Indirect Methods for Estimating Environmental Benefits." American Journal of Agricultural Economics 28 (1986): 280-290.

Smith, V. Kerry, William H. Desvousges, and Ann Fisher. "Estimating Environmental Benefits." American Journal of Agricultural Economics 68 2 (May 1986).

Sudman, S. and N. M. Bradburn. Response Effects in Surveys. Chicago: Aiding Publishing Company, 1974.

Tversky, Amos, and Daniel Kahneman. "Judgement Under Uncertainty: Heuristics and Biases." Science 185 (September 27, 1974): 1124-1131.

Wyner, Gordon A. "Response Errors in Self-Reported Number of Arrests." Sociological Methods and Research 9 2 (November 1980): 161-177.