WATER RESOURCES RESEARCH, VOL. 29, NO. 7, PAGES 2445–2454, JULY 1993

342

The Value of Clean Water: The Public's Willingness to Pay for Boatable, Fishable, and Swimmable Quality Water

RICHARD T. CARSON

Department of Economics, University of California, San Diego

ROBERT CAMERON MITCHELL

Graduate School of Geography, Clark University, Worcester, Massachusetts

This paper presents the findings of a study designed to determine the national benefits of freshwater pollution control. By using data from a national contingent valuation survey, we estimate the aggregate benefits of meeting the goals of the Clean Water Act. A valuation function is estimated which depicts willingness to pay as a function of water quality, income, and other variables. Several validation checks and tests for specific biases are performed, and the benefit estimates are corrected for missing and invalid responses. The two major policy implications from our work are that the benefits and costs of water pollution control efforts are roughly equal and that many of the new policy actions necessary to ensure that all water bodies reach at least a swimmable quality level will not have positive net benefits.

INTRODUCTION

In 1972 Congress passed the Clean Water Act (PL 92-500) by an overwhelming margin and thereby declared a national goal of zero discharge of pollutants into navigable waters by 1985, an interim goal of achieving waters fit for fishing and swimming by 1983, and the requirement that water pollution sources meet nationally uniform pollution control technology standards established by the U.S. Environmental Protection Agency (EPA). During the debate over this Act, Congress paid little attention to the cost of achieving these ambitious goals because the benefits seemed large and the costs were largely unknown and spread through time [*Leone and Jackson*, 1981]. It was anticipated that Congress could correct the Clean Water Act on the basis of future research and experience. Since 1983, however, despite the failure to meet the 1983 and 1985 goals, attempts to make corrections have foundered due to disagreement over what should be done. In 1987, Congress postponed compliance deadlines for effluent standards and put into place requirements for new state programs to deal with nonpoint sources. The Clean Water Act will again be up for renewal in 1993. Given the latest Department of Commerce estimate [*Bratton and Rutledge*, 1990] of current annual spending on controlling freshwater pollution, $37.3 billion (1990 dollars) in 1988, and the increase in expenditures necessary to achieve further improvements, it is important to assess the aggregate economic benefits of attaining the fishable-swimmable goal.

Those who would estimate the economic benefits of national water pollution control programs face a dilemma. On the one hand, data at the national level have thus far been lacking or are untrustworthy [*Dorfman*, 1977; *Feenberg and Mills*, 1980; *Freeman*, 1982]. On the other hand, studies that have valued "local" water bodies such as river basins or lakes, while more numerous [e.g., *Greenley et al.*, 1982] are, at best, of limited use in determining the benefits of national water quality policy changes. Theoretical work by *Hoehn and Randall* [1989] demonstrates that independently derived benefit estimates for geographic locations or categories of benefits which are potential substitutes or complements for each other can not be aggregated to obtain national benefits in a straightforward manner. *Hoehn* [1991] provides additional theoretical discussion on this issue and provides empirical evidence that performing such an aggregation may result in a significant overestimate of total benefits.

The approach presented here avoids the geographical aggregation problem by the use of a contingent valuation (CV) study in which a national probability sample of American households was asked to value a national set of water quality improvements. It avoids the problem of benefit category (e.g., fishing aesthetics) aggregation by asking the same sample to give the total value of all benefits for them for three progressively higher levels of water quality. In addition to directly measuring national water quality benefits, the national CV approach also makes it possible to estimate a valuation function which predicts willingness to pay as a function of the level of water quality, income, water-based recreational use, and environmental attitudes and which we use to update our 1983 benefit estimates to 1990 dollars. Other features of the study include a design that allowed us to test for part-whole bias and reconsideration effects and the use of a new missing data imputation method to compensate for the bias introduced by missing valuation responses. Of particular interest, given the strong claims made by authors such as *Kahneman and Knetsch* [1992] that the embedding effect is "the most serious shortcoming of CV," is the comparison between our national water quality benefit estimates and those made by *Smith and Desvousges* [1986] for the same quality changes in a subnational resource: the Monongahela River.

Paper number 93WR00495.
0043-1397/93/93WR-00495$05.00

CONTINGENT VALUATION

The contingent valuation method uses survey research techniques to elicit people's preferences in the form of willingness-to-pay (WTP) monetary amounts. In its standard form, the CV survey describes a detailed hypothetical market in which a specified good may be purchased and asks respondents how much of their current household income in dollars they would be willing to give up in exchange for a specified increase in the level of the public good. Thus we ask the respondent for a direct evaluation of his household's compensating surplus (CS) from a change in the good in question which can be represented as

$$CS = [e(p_0, q_*, q_0, U_0)] - [e(p_0, q_*, q_1, U_0)], \quad (1)$$

or equivalently, $CS = Y_0 - Y_1$, where e is the expenditure function, p_0 the vector of prices for marketed goods, q_* is the vector of nonmarketed goods which remain fixed, q_0 and q_1 are the initial and subsequent levels of the nonmarketed good being valued, respectively, Y_0 and Y_1 are the initial and subsequent levels of disposable income associated with each of the two expenditure functions, respectively, and U_0 is the initial utility level. An equivalent representation of (1) which is often more useful for estimation purpose is the income compensation function [*Willig*, 1976] $\mu(q_1|p_0, q_*, Y_0, V(p_0, q_*, q_0, Y_0, T))$, where T is a vector of taste parameters and V is the indirect utility function which is assumed here to equal the utility level U_0. Often the valuation question is repeated several times for different levels of the good so that a Hicksian compensated demand curve can be traced out.

Since its initial applications in the 1960s, considerable effort has been devoted to establishing its theoretical basis, developing the actual methodology and, where possible, comparing its estimates with those using market demand based measures [*Cummings et al.*, 1986; *Mitchell and Carson*, 1989]. Even though the method is now the one most frequently used to value environmental amenities (*Carson et al.* [1993] provide a bibliography of over 1200 papers on contingent valuation); its use has its critics [e.g., *Cambridge Economics*, 1992]. If one accepts the compensating surplus form of WTP as the appropriate welfare measure for a specified improvement in the water quality enjoyed by an individual household and takes the current distribution of income as given, then a point on the Samuelson-Bradford bid/benefit curve [*Randall et al.*, 1974] is given by summing all household's WTP amounts for the new level of water quality. Optimal provision of water quality occurs at the point where the aggregate marginal cost and marginal benefit curves cross.

Table 1 shows the types of benefits which might accrue from improved water quality. In the presence of uncertainty, contingent valuation obtains an estimate of the difference between two planned expenditure functions, an ex ante welfare measure which is referred to as option price [*Smith*, 1987]. Several benefit categories, particularly those making up the nonuse (i.e., existence) class of benefits, may show no traces in marketplace transactions. There are reasons to believe that this class of benefits may comprise a sizable portion of total water quality benefits [*Fisher and Raucher*, 1984]. Among benefit measurement techniques, the contingent valuation method is uniquely able to measure nonuse benefits since it can elicit values from both users and nonusers of a given amenity.

TABLE 1. A Typology of Benefits From an Improvement in Freshwater Quality

Benefit Class	Benefit Category	Benefit Subcategory (Examples)
Use	In stream	Recreational (water skiing, fishing, swimming, boating)
		Commerical (fishing, navigation)
	Withdrawal	Municipal (drinking water, waste disposal)
		Agriculture (irrigation)
		Industrial/commerical (process treatment, waste disposal)
	Aesthetic	Enhanced near water recreation (hiking, picnicking, photography)
		Enhanced routine viewing (commuting, office/home views)
	Ecosystem	Enhanced recreation support (duck hunting)
		Enhanced general ecosystem support (food chain)
Nonuse	Vicarious consumption	Significant others (relatives, close friends)
		Diffuse others (American public)
	Stewardship	Inherent (preserving remote wetlands)
		Bequest (family, future generations)

DESIGN CONSIDERATIONS AND FEATURES

Although there is a growing consensus that the CV method is a valid way of measuring the value of nonmarket goods, successful implementation of the method requires great care because survey-based benefit estimates are sensitive to methodological artifacts. Large errors are possible if the survey instrument is misunderstood or not accepted as plausible by respondents, or if the questions or the interviewers improperly influence the respondents' answers, or if the sample is distorted in some way. In this section we provide information about the relevant design features of our instrument which were developed after an extensive period of questionnaire design research and pretesting. The complete CV survey instrument is contained as an appendix in the work by *Mitchell and Carson* [1989]. It was developed and tested in three phases. During the first phase, 1979–80, we developed a precursor to the present instrument which we tested by appending it to a national, in-person survey of 1576 respondents we were conducting for another study [*Mitchell and Carson*, 1981]. We used equivalent subsamples in this pilot study to test the payment card elicitation method for bias. In the second phase, we collaborated with the Research Triangle Institute in the summer of 1983 to pretest a modified and expanded version of this instrument. This led to further modifications to make it more comprehensible to respondents. Finally, just prior to the survey's administration in November 1983, further pretesting was conducted by the Opinion Research Corporation to ensure that the survey was suitable for administration by their national field staff. Additional information about the instrument, its administration, the statistical tests reported here, and other issues related to the valuation of national water quality improvements may be found in our three technical reports on this project [*Mitchell and Carson*, 1981; *Mitchell and Carson*, 1984; *Mitchell and Carson*, 1986].

One issue is whether, compared with local or regional water quality, national water quality is too far removed from the respondents' everyday life and experience for them to find the valuation scenario a plausible choice situation. Plausibility was enhanced by several factors. One factor was the all-inclusive nature of the amenity valued. Pretests showed that respondents quickly grasped the idea of improving the water quality of all lakes, rivers, and streams throughout the country. The plausibility of valuing national water quality was also enhanced by the widespread awareness among the respondents that water quality policy is made at the national level. It was necessary, of course, to define the amenity change in specific terms. Here plausibility was aided by the fact that the levels of minimum national water quality we asked respondents to value, boatable, fishable, and swimmable, are concepts that are widely understood (although not unambiguously interpreted) in our society. That they are also mandated by the Clean Water Act enhances the correspondence between our valuation questions and the behavior we seek to predict. This does mean, however, that our results will be more useful in making broad comparisons of the benefits and costs of different national water quality objectives than in making detailed benefits estimates for the types of control problems routinely faced by government regulators.

Matching the boatable, fishable, and swimmable levels of water quality with physical water quality criteria is no easy task, nor is there complete agreement on how to do this. In our survey instrument, we used the Resources for the Future water quality index developed by W. J. Vaughan for our pilot study which maps the rungs of the water quality ladder back into physical water quality parameters such as dissolved oxygen which can in turn be tied to a large scale water quality model. Use of this ladder in the survey as a visual aid greatly facilitated the task of communicating the several quality levels to the respondents. *Smith and Desvousges* [1986] successfully used the ladder for the same purpose in their Monongahela River Study.

The scenario's wording emphasized the nonuniform distribution of water quality implied by the concept of "minimum" water quality. Respondents were told that although the present minimum level is boatable, most of the nation's freshwater bodies are currently fishable and perhaps 70–80% are swimmable. When asked to value the boatable (minimum) level, respondents were asked how much they would be willing to pay "to keep the nation's freshwater bodies from falling below the boatable (minimum) level where they are now." This "below boatable" baseline, which was described in some detail in the scenario, represented the minimum level of national water quality which would occur if all present annual expenditures for water pollution control by industry and governmental entities ceased. The U.S. Commerce Department also measures water pollution control expenditures from this baseline. This baseline offered the respondents the opportunity to purchase water quality improvements in a form which allows the compensating surplus-WTP measure to be used in all the valuation questions.

The payment vehicle used in this study was annual taxes and higher product prices. While contingent valuation estimates are not generally independent of the payment vehicle used, the one used here has the advantages that it corresponds with the way citizens presently pay for water quality and that it was accepted by most respondents without protest as appropriate for this purpose. In an effort to avoid the starting point bias associated with the commonly used bidding game method [*Boyle et al.*, 1985; *Randall et al.*, 1974], our elicitation procedure used a grounded payment card format we first developed and tested in the pilot study for this project. Respondents were divided into five income groups and given a payment card containing a large selection of amounts ranging in order from $0 to a very high dollar amount. In order to provide a meaningful context for the valuation exercise, five points on the continuum were identified as the amounts average households of that income group are currently paying in taxes and higher prices for nonenvironmental public goods such as defense, the space program, and police and fire protection. The willingness-to-pay questions asked respondents to state an amount on the payment card or "any amount in between" they were willing to pay for each of the three levels of national minimum water quality.

Because our development work indicated that some respondents tend to confuse drinking water benefits with freshwater benefits, the scenario was worded to distinguish the two types of benefits. To ensure that all respondents were aware of the full range of appropriate benefits, a "values" card [*Smith and Desvousges*, 1986] was used which listed the major reasons why households might value water quality. We did not intend the respondents to take any of the commercial in-stream or withdrawal benefits described in Table 1 into account and it is unlikely, given the wording of the CV scenario, that they did so. To promote respondent understanding of the water quality levels, additional descriptions were provided. Regarding the fishable level, for example, respondents were told that "although some kinds of fish can live in boatable water, it is only when water gets this clean that game fish like bass can live in it." The scenario also reminded respondents that they are currently spending part of their income on water pollution control, a condition they needed to understand for us to implement our WTP-compensating surplus questions. Our pretests had found that a number of respondents wanted to know how much they were paying for this purpose. This created a potential problem since we could not inform them of this amount prior to eliciting their WTP amounts because of the likelihood that some would base their value on this figure instead of independently determining their maximum WTP amount. By offering to provide this information at a later stage in their interview and give the respondents an opportunity to revise their original WTP amounts on the basis of this information if they wished, we were able to coax reluctant respondents to give us initial values and later to test the effect of providing this information.

During the course of the interview we solicited a total of four WTP amounts from each respondent for each of the three water quality levels in order to provide opportunities for respondents to arrive at a considered value. The first (WTP_F) series consists of the amount given for each of the three WTP questions (boatable, fishable, and swimmable). The corrected (WTP_C) series consists of the amounts (whether changed or unchanged) offered after each of their first amounts were repeated to the them, the total was stated, and they were encouraged to make any corrections they wished. The informed (WTP_I) series consists of the amounts given after respondents were informed of the range of the amounts households in their income group were actually paying for water (and air) quality. Finally, the pushed (WTP_P) series consists of amounts obtained by pushing the respondents to increase their WTP amounts by telling them that the amounts they previously stated were not enough to reach any of the three goals, including the boatable water quality goal.

Given the variety of measures, to three quality levels by four separate measurements, it is at this point useful to make clear our assumptions about the nature of the amounts elicited. We assume that many or most of the respondents do not have a well formed value when asked in a CV survey to value a good which they are unaccustomed to purchasing. Faced with such a first-time request for such a value, some respondents are unable to offer a value. The interviewers were instructed to avoid putting pressure on these respondents to give what would only amount to meaningless values. The remaining respondents, however, know within a reasonable range where their value for the good may lie, and a few may even have a good idea of the actual value. On the assumption [*Hoehn and Randall*, 1987] that respondents are generally cautious (i.e., they are risk averse consumers) when faced with sizable purchases, we believe the MIN (WTP_C, WTP_I) represents the likely lower bound of their WTP range. In the case of the WTP_P amounts, the request for revaluation is likely to have been interpreted by some respondents as implying that they had not given a high enough WTP amount and "should" give more. The prospect offered respondents in this scenario was quite drastic, suggesting that even the boatable level was threatened if a higher WTPB bid was not forthcoming. On the assumption that respondents can be pushed by the interviewers to give amounts at or above their maximal willingness to pay, the WTP_P amounts represent an upper bound on the desired WTP measure. The difference between the $WTP_P$and WTP_I series may be indicative of how far respondents can be encouraged to revise their WTP amounts upward by interviewer and social pressure.

A major question in valuing water quality improvements is the shape of the benefit curve between the three goals of boatable, fishable, and swimmable water quality levels. If willingness to pay is completely contingent upon the attainment of each goal, the function is a step function, and intermediate or partial improvements would provide no additional benefits. To examine this possibility, we asked a "halfway" policy question of half the sample. These respondents were asked if they would still be willing to pay their corrected amount for swimmable "if the best we could do was to raise the minimum only halfway from fishable to swimmable." The overall improvement conveyed to the respondent in this treatment was substantially more than halfway, as respondents were told "At halfway, more water bodies would be improved over the fishable level, and some additional, but not all, water bodies would be improved to the swimmable level."

Another direction in which the willingness to pay for water quality improvements response surface can be explored is geographical. One way we did this was by asking a 95% question of the other half of the sample. This question asked respondents if they would still be willing to pay the fishable amount if "five percent of the nation's water bodies remain at the boatable level .... The lakes, rivers, and streams comprising this five percent would all be located in heavily industrial and/or urban locations where a lot of people live." A second way is to look at the extent to which respondents value provision of the good outside their home area. Pretests showed the most readily understood definition of a home area for a survey such as ours was the respondent's state. Any definition of "local" area is necessarily ad hoc. The state definition may have some relevancy in its own right due to an increasing emphasis upon delegating authority for environmental standards and enforcement to the states. We asked respondents how many dollars or what percentage of their $WTPTOT_C$ bid they would give to their state and to the rest of the nation.

It is sometimes believed that a respondent who is asked a value for a particular water quality level unwittingly values, instead, a more general package of environmental improvements. To minimize the possibility of this type of bias, termed policy-package-part-whole bias [*Mitchell and Carson*, 1989], we explicitly asked respondents to keep in mind that no matter what amount they give for water pollution control, they will also continue to pay for the nation's other environmental programs such as air pollution and "air quality will remain at its present level or improve slightly." To test for the presence of this bias, we again used a split sample approach. At the point in the interview where respondents were told what amount they are already paying for pollution control, those in subsample A were told only the amount for water quality control whereas respondents receiving treatment B were given the amounts for both water and air pollution control. If this type of part-whole bias was present in the WTP_C bids, we hypothesize that the provision of the air pollution cost information to subsample B respondents would cause that subsample to disproportionately reduce their WTP_I amounts to compensate for their previous overspending of their environmental account.

A national area probability sampling plan based on the 1980 Census was used in this study. Experienced professional interviewers under the supervision of the Opinion Research Corporation conducted in-person interviews at 61 primary sampling points in the contiguous United States. Each interview took approximately 40 min. The response rate was 79% of eligible respondents and a total of 813 people were interviewed. Most of those counted as nonrespondents were households never found at home in spite of repeated call backs. Complete details of the sampling plan and its execution can be found in the work by *Mitchell and Carson* [1984].

FINDINGS

Of the original 813 interviews, 564, or 70% yielded "usable" WTP amounts. The remainder consisted of 72 don't knows (9%), 18 refusals to answer the WTP questions (2%), 133 protest zeros (17%), 16 inconsistent (too high) responses (2%), and 10 inconsistent (too low) responses (1%). Responses judged inconsistent (too high) were those which exceeded 5% of the household's income, while those judged inconsistent (too low) were WTP amounts of less than $5.00 (usually $1.00) given by respondents with above average to high incomes whose answers to attitude questions showed strong support for water pollution control expenditures. Given the degree of interest and the effort involved in

TABLE 2. Mean Unadjusted Annual Household Willingness to Pay Amounts for Different Levels of National Water Quality by Type of Bid

Water Quality Level	First Bid (*F*)	Corrected Bid (*C*)	Informed Bid (*I*)	Pushed Bid (*P*)
Nonboatable to boatable (WTPB)	$111 (10;$40)*	$106 (10;$40)	$125 (11;$48)	$141 (13;$50)
Boatable to fishable (WTPF)	$80 (8;$30)	$80 (8;$30)	$96 (9;$35)	$108 (10;$50)
Fishable to swimmable (WTPS)	$89 (12;$25)	$89 (12;$25)	$102 (12;$25)	$116 (13;$25)
Total WTP (WTPTOT)	$280 (25;$125)	$275 (25;$120)	$323 (27;$150)	$366 (29;$150)
Number changing their bids at each stage		75	104	136

$N = 564$.
*Standard error of the mean; median.

answering complex CV scenarios such as the one used in this study, this level of item response in a national sample, while high, may be acceptable, provided the estimates are adjusted to take into account the fact that the nonrespondents were not a random subset of the sample. These adjustments are discussed later in the paper.

Table 2 presents the unadjusted WTP amounts for each of the four series of bids measured in the study. Using the corrected series of bids, the respondents who gave usable responses were willing to pay on average $106 annually for maintaining boatable quality water ($WTPB_C$), $80 more to reach the fishable minimum water quality level ($WTPF_C$), and an additional $89 to move from the fishable minimum quality to a national minimum of swimmable quality water ($WTPS_C$) for an unadjusted mean total ($WTPTOT_C$) of $275. An examination of the changes made by the 75 respondents who reconsidered and corrected their amounts after giving their WTP_F amounts shows that most of them corrected mistakes caused by misconceptions about the elicitation process. In the WTP_I iteration, 104, or 18% of the respondents corrected their bids after being informed of the approximate level of their current payments for water quality (water and air quality) improvements.

Those changing bids tended to be respondents who discovered that they were actually paying more money than their previous (corrected) amount and wished to increase their WTP_C amounts. Of those who discovered that they were actually paying less than they said they were willing to pay, few reduced their earlier bid. Those who changed their amounts in the last iteration (WTP_P), after being confronted with the assertion that the amount they had previously committed themselves to might not be enough to maintain even the present minimum level of water quality (a strong statement), tended to be of two types: (1) respondents whose informed bid was still below their current payments and (2) respondents who already had a corrected bid much higher than their current payments. Overall, taking those who made multiple changes into account, approximately 30% of the respondents changed one or more of their WTP amounts. Of these respondents about a third changed more than once. The mean bids for the first and corrected conditions were not significantly different based on t tests computed using a test-retest framework. Future opportunities to change generally resulted in mean bids that were insignificantly higher than their predecessors, although the WTP_C series are all significantly lower than the WTP_P series. We believe the WTP_C series represents the most valid basis for estimating WTP (after adjustment for nonresponse). The informed series and, in particular, the highest series of WTP questions put a significant amount of pressure on the respondents to increase their willingness to pay, and should be viewed as upper bounds. The WTP_P series is 30–35% larger than the WTP_C series.

The results of the test to see if some respondents had believed they were valuing a broader pollution control policy which includes air pollution control are reassuring. None of the t tests of the differences in the willingness to pay for any water quality level between subsample A (who were told what they were paying for water quality) and subsample B (who were told the amounts they are currently paying for both air and water quality improvements) has an absolute value greater than 0.75. This indicates that we can reject the hypothesis that providing information on air pollution control costs to subsample B caused that subsample to disproportionately reduce their WTP amounts. In turn, this provides support for the proposition that a particular type of policy-package part-whole bias, erroneous inclusion of air quality values, is not a major problem for this study.

A Valuation Function

A total value/bid curve for the WTP_C can be specified in terms of the Hicksian income compensation function whose arguments are a base water quality level q_0; the level of water quality being valued q_i; disposable household income Y_0; and the taste variables of water-based recreational use W_r, and environmental attitudes, A_e:

$$\text{TOTWTP}_i = f(q_i, Y_0, W_r, A_e | q_0). \tag{2}$$

Differentiating this bid curve with respect to q_i, $\partial\text{TOTWTP}_i/\partial q_i$ yields the inverse Hicksian compensated demand curve.

Stacking the observations from each of the quality levels which were specifically asked about and using a Cobb-Douglas representation of the income compensating function which fits substantially better than the linear representation, we obtain the following results:

$$\begin{aligned} \text{TOTWTP}_i = \exp\,[&0.413 + 0.819^* \log (q_i) \\ &(1.66) \quad (9.20) \\ &+ 0.959^* \log (Y_0) + 0.207^* W_r + 0.460^* A_e], \quad (3) \\ &(22.30) \qquad\qquad (3.23) \qquad (6.47) \end{aligned}$$

where q_i is the numeric value (e.g., boatable) on the water quality ladder being valued, Y_0 is the current annual household income in thousands of dollars and proxies for the desired disposable income, W_r is a dummy shift variable for whether or not any member of the household engaged in freshwater boating, fishing, or swimming activity during the previous year, and A_e is a dummy variable for whether or not the respondent regarded a national goal "of protecting nature and controlling pollution" as "very important." For the sample (n = 1599) on which this regression was based, Y_0 = 24.22, W_r = 0.59, and $\bar{A}_e$ = 0.65. This Cobb-Douglas model with an additive error term [*Goldfeld and Quandt*, 1972] predicts mean WTP conditional on the covariates and can be estimated using nonlinear regression techniques. Note that if (3) is estimated in its linear form after taking logarithms, as is most often done in a Cobb-Douglas model, then one is estimating the conditional median WTP rather than mean WTP [*Goldberger*, 1968]. The t statistics given in the parentheses are based on the heteroskedasticity consistent covariance matrix proposed by *White* [1980]. The coefficients in this equation are all reasonable in terms of sign and magnitude and are all quite significant. The adjusted R^2 for this equation, 0.27, is large relative to many contingent valuation surveys, especially given the small number of variables in the valuation function. Equation (3) forecasts well, at least within sample. For example, the WTP amount predicted for a boatable to fishable water quality change using (3) evaluated at the mean values of Y_0, W_r, and A_e is $79, compared to the observed sample mean of $80 for the same change.

PARTIAL IMPROVEMENTS

According to the answers to the "halfway" and the "95 percent" questions, the benefits of partial improvements are considerable. For the halfway improvement question which was asked of subsample A, we find that 73% of the respondents said that going halfway from a fishable to a swimmable minimum national water quality was worth the same to them as the total improvement to the swimmable level. It is important to note that this halfway improvement would have provided a substantial amount of swimmable quality water and, as a result, is likely to be perceived as providing for a non-marginal increase in the quantity of swimmable quantity water. Quoting from the text read to the respondents, "At halfway [between a minimum fishable and swimmable quality level] . . . , some additional but not all water bodies would be improved to the swimmable level." The $84 WTP amount for the halfway improvement is 6% less than the $89 the respondents were willing to pay for the swimmable level (WTPS_C).

Turning now to the 95% question, almost nine out of ten (89%) of those who were asked this question (subsample B) said this amount of improvement from boatable to fishable was worth the same to them as a complete improvement (defined as where 99% or virtually all the nation's lakes, streams, and rivers would be fishable). Those who wished to pay less for the partial improvement were disproportionately residents of large urban areas. This is understandable because the question informed respondents that the "lakes, rivers, and streams comprising this five percent would all be located in heavily industrial and/or urban locations where a lot of people live." Each person who was unwilling to pay the same amount was asked how much he or she was willing to pay for this partial improvement. Because those who were not willing to pay the same amount were willing to pay a somewhat lower amount for the improvement than in the halfway question, the WTP amount for raising 95% of the nation's water to at least the fishable level is $74, or 8% less than the $80 amount ($\text{WTPF}_C$) for raising 99% to at least this level. Both the 95% and the halfway questions attempt to get at the same fundamental question, Do benefits fall dramatically if one backs away slightly from a uniform water quality goal? Our results suggest they do not, although substantially more research is needed to determine how benefits change with small changes in water quality and, in particular, the spatial location of those changes.

The other question which relates to the spatial nature of improvements asked the respondents how they would divide their WTPTOT_C between their state and the rest of the nation. They allocated an average of 67% of this amount to be spent in their state and 33% for out-of-state improvements. While these data show significant in-state benefits, the level of out-of-state benefits is consistent with a strong federal role in water pollution control.

A TEST OF EMBEDDING

One important issue in using surveys to value national environmental programs is whether respondents can meaningfully value amenities at this level of abstraction in a meaningful way. We examine this issue here in the context of a test of the embedding proposition. *Mitchell and Carson* [1986] examine it in the context of a benefit-transfer exercise.

The embedding proposition was first put forth by *Kahneman* [1986], who found respondents in a telephone survey were willing to pay essentially the same amount for all lakes in Ontario as for lakes in a small area of Ontario. *Kahneman and Knetsch* [1992] provide another example of this phenomena, this time using more and less inclusive policies, and label the embedding problem "perhaps the most serious shortcoming of CVM." *Smith* [1992] in a critique of Kahneman and Knetsch's paper argues that their results may be largely due to bad survey design, vaguely described goods.

It is important to note an important distinction between part-whole bias as defined by *Mitchell and Carson* [1989] and the embedding problem as considered by *Kahneman and Knetsch* [1992]. Part-whole bias occurs when a respondent values a larger or smaller entity (e.g., geographic location, range of benefits, policy package) than that intended by the researcher. Part-whole bias is a common but avoidable obstacle in designing a contingent valuation survey, and its avoidance is seen by *Mitchell and Carson* [1989] as a key element in the survey designer's primary task, ensuring that the respondent is valuing the good intended

In contrast, *Kahneman and Knetsch* [1992] seem to see embedding as an indicator of the inevitable inability of respondents to rationally value public goods. It will be useful to state their definition of embedding: "the same good is assigned a lower value if WTP for it is inferred from WTP for a more inclusive good than if the particular good is evaluated on its own." This definition is problematic because the sequence in which goods are valued, substitution effects, income effects, and changes in the composition of the choice set should make a difference in the agent's valuation. Thus embedding can only be seen as a distinct problem with contingent valuation if this so-called embedding effect is much larger than that which can be explained by reference to plausible economic effects, that is to say that the contingent valuation method produces the effect while it is not seen in the actual behavior of consumers or voters. Because it is hard to judge what the exact magnitude of these economic effects should be, a statistical test of either part-whole bias or embedding will set out as its null hypothesis that two related goods (where one encompasses the other) are valued identically by respondents against the alternative that the good which encompasses the other is valued more highly. Note however, the *Mitchell and Carson* [1989] framework suggests that while the null hypothesis may be accepted in a poorly designed contingent valuation study, it should be rejected by a well-designed contingent valuation survey. The Kahneman and Knetsch framework, because it views the phenomena as "inevitable," suggests that the null hypothesis should always be accepted. The difference between the two frameworks is that accepting a null hypothesis under Mitchell and Carson framework implies a rejection of a particular study while under the Kahneman and Knetsch framework it implies rejection of the entire contingent valuation method.

The availability of a CV study [*Smith and Desvousges*, 1986] for a regional freshwater resource, the Monongahela River Basin in western Pennsylvania, whose design is comparable to ours in several important respects, offers an opportunity for an empirical assessment of this issue. It is possible to conduct an almost ideal test of geographic part-whole/embedding hypothesis using our data set and that of *Smith and Desvousges* [1986]. The ideal test would involve two large independent in-person surveys of the same population with highly trained interviewers, conducted over the same time period, using survey instruments which differed only in the specific details of the two goods being described, the two goods having well-defined relationship to each other, one inclusive of the other, and the "difference" between the two goods was such that one would expect, on economic grounds, to see a fairly large difference in the respondents' valuation of them. A comparison of our national results and Smith and Desvousges' Monongahela results meets all of these criteria except the need for two random samples of the same population at the same time: ours is a national sample theirs is a sample of the Pittsburgh area, a little less than 2 years separating the interview periods. Both of these weaknesses are easily correctable. We can make a small CPI adjustment to Smith and Desvousges' estimates to correct for the time difference, and we can evaluate our valuation function at Smith and Desvousges' covariate means to get an estimate of what a Pittsburgh sample would pay for a national water quality change.

Evaluating the function for the boatable to fishable quality water change, which we will use as an example, yields an estimate of $68, which is $11 less than our estimate for the national sample. Now we need to statistically test the difference between our estimate, $68, and *Smith and Desvousges*' [1986] estimate of $26 for their sample's average willingness to pay for the same change on only the Monongahela River. The simplest test is to use our estimate ($68), our unconditional standard deviation ($92), and sample size ($n = 564$) for our boatable to fishable quality change and compare it to Smith and Desvousges' estimate ($26) and standard deviation ($39) from Smith and Desvousges' combined sample ($n = 211$). The resulting approximate t test based on unequal variances is 4.88 and has a p value of less than 0.001. Thus the part-whole hypothesis in this case is rejected. More powerful tests can be created by conditioning on the potential covariates in the two samples. This substantially reduces the (unexplained) variance of the WTP amounts in the two samples and hence result in much larger t statistics. We also get stronger rejections with comparisons based on the other water quality levels. It is important to keep in mind that while the results in this section suggest that our respondents and those of Smith and Desvousges valued different goods differently, a test such as this one can not conclusively demonstrate that either set of respondents valued the good intended by the researchers.

COMPARISON WITH EARLIER CV STUDIES

Another issue is the stability of CV estimates over time and across CV instruments. *Gramlich*'s [1977] 1973 CV study of Boston Residents' WTP for water quality improvements in the Charles River offers an opportunity to investigate this issue because he also asked his respondents how much they were wiling to pay for swimmable quality water throughout the United States from a baseline level of water quality which was similar to our boatable or perhaps a little above. Inflated to 1983 dollars, his mean WTP estimate of $55 represents a value of $130 for this change. Given the differences in samples, methods, and time periods, this amount is reasonably close to our $169 for roughly the same change. Indeed, the difference is not significant at the 10% level using a t test.

A more direct comparison to our estimate of the value of achieving a national level of swimmable quality water can be obtained by looking at our 1980 pilot study [*Mitchell and Carson*, 1981], which used a large sample ($N = 773$). Updating that 1980 number to 1983 using the CPI shows the mean estimate for swimmable quality water ($275) from the 1980 survey to be almost identical to the WTP_C mean estimate from the 1983 survey. This comparison provides important evidence on the replicability of a contingent valuation survey instrument.

CORRECTION FOR MISSING AND INVALID VALUES

In order to obtain an aggregate estimate of the benefits of freshwater quality improvements, it is necessary to generalize from our sample to the national population. One potentially serious problem in contingent valuation surveys is the need to adjust the data to compensate for the bias introduced by the inevitable failure to interview every person selected for the sample (unit nonresponse) and by the failure of some

TABLE 3. Adjusted Annual Household Values for Best Estimate of National Water Quality Benefits

	Mean	Standard Error of the Mean	95% Confidence Interval
WTP_c (boatable)	\$93	\$8	\$77–109
WTP_c (fishable)	\$70	\$6	\$58–82
WTP_c (swimmable)	\$78	\$9	\$60–96
$WTPTOT_c$	\$242	\$19	\$205–279

respondents to give valid answers to the WTP questions (item nonresponse). Procedures to compensate for these two types of nonresponse are routinely used by the Census Bureau and other government agencies [*Madow et al.*, 1983]. Sample nonresponse is usually corrected by some type of weighting procedure, while item nonresponse is usually corrected by some type of imputation procedure. Item nonresponse, when defined in a broad sense to include various types of invalid or unusable responses such as protest zeros as well as "don't know's," is a particularly troublesome problem in CV surveys such as ours which use comparatively complicated scenarios.

We adjusted the data from the WTP_C (corrected) series to compensate for item and unit nonresponse in two steps. First, the WTP values for the 30% of the respondents with missing or invalid WTP values were imputed using CART, a powerful nonparametric tree-structured classification procedure proposed by *Breiman et al.* [1984]. CART searches over all values of available variables to determine the binary split (i.e., constructed dummy variable) which best predicts the variable of interest. This process is repeated using a series of nodes and branches with the data at each node being split into separate subsamples on the basis of the optimal binary predictor at that node. Cross-validation is used to "prune" the tree grown to prevent overfitting.

Second, we corrected for sample nonresponse by using the household weights supplied by the Opinion Research Corporation to weight the observations to make the available sample (i.e., valid plus imputed responses) more representative of the Census population. As is typical in completed national probability sample surveys, women are somewhat overrepresented in our unweighted sample of respondents, and young black males are underrepresented. A combination of household weights and imputing the missing values reduces the adjusted $WTPTOT_C$ value by 12% with each of the two correction techniques contributing approximately equally to this reduction. This scale factor was applied for consistency to the rest of the WTP_C series as shown in Table 3.

AGGREGATE BENEFITS

In this section, we take benefit estimates from our own work and that of others from the early 1980s and update those estimates by correcting for the CPI and the current number of national households. For the benefits categories measured in our survey, our best estimate of the benefits of achieving the national swimmable water quality goal from a baseline of nonboatable water is \$29.2 billion dollars a year (1990 dollars). Extreme bounds of \$24 to \$45 billion dollars a year in swimmable benefits can be developed by taking the lower 95% confidence interval for $WTPTOT_C$ and the upper 95% confidence interval for $WTPTOT_P$. Because the $WTPTOT_P$ elicitation questions deliberately pressured respondents to change their amounts upward, however, a more reasonable range of benefits for the swimmable quality goal is \$24 to \$40 billion dollars a year.

It is of interest to compare our national estimate with the estimate *Freeman* [1982] developed on the basis of a review of the available studies, each of which had estimated the national benefits associated with one or two water related activities (such as fishing), a methodology that requires very strong separability assumptions. This comparison is complicated, however, by an important difference between his goal and ours: Freeman sought to measure the benefits of going from the 1972 water quality level to achieving the ambient quality levels believed to be associated with meeting best available technology (BAT) provisions of the Clean Water Act, whereas our goal is to measure the benefits of going from no pollution control to achieving swimmable quality water nationwide. Freeman's aggregate estimate also includes two categories of benefits, commercial usage and marine recreation, which were not covered by our CV survey. In 1990 dollars, Freeman's point estimate is \$20.1 billion dollars with a range of \$8.2 and \$39.6 billion dollars. In addition to a CPI adjustment, we have also adjusted Freeman's estimates upward by the percentage increase in the number of households from 1980 to 1990. If we add his estimate of \$9.9 billion for commercial usage and marine recreation benefits to our preferred point estimate, our annual benefit estimate for all categories of benefits increases to \$39.1 billion for attainment of the swimmable water quality goal. That our estimate is higher than Freeman's is to be expected given the larger magnitude of the water quality improvement measured in our study.

While we have adjusted both sets of aggregate estimates for changes in population and the consumer price index, we have not adjusted them for changes in the variables which the valuation function above suggest determine a household's willingness to pay for water quality improvements: income, water-based recreation, and attitudes toward pollution control. Exact comparisons between the 1983 values and 1990 values for these variables are not available. However, Census Bureau information suggests an approximately 10% increase in "real" household income. The University of Chicago's National Opinion Research Center's General Social Survey suggests an approximate 30% increase in the number of respondents who think that there should be more spending on pollution control, and other survey organizations, such as Roper, report similar or larger changes in public attitudes toward the environment and water pollution in particular. The large percentage of households engaged in water-based recreation appears not to have substantially changed. Incorporating a 10% increase in real income (Y_0) and a 30% increase in attitudes toward pollution (A_e) results in a 19.5% increase in willingness to pay for swimmable quality water. Scaling the \$39.1 billion up by this factor results in a \$46.7 billion dollar estimate.

How do the aggregate benefits implied by our data compare with the current and projected costs of all water pollution control programs? This comparison, if we ignore the toxics issue, is relatively straightforward because our measure of benefits is directly tied to ambient water quality levels and the Commerce Department cost estimates include all expenditures for water pollution control irrespective of whether they were taken in response to particular provisions of federal legislation. The Department of Commerce estimates [*Bratton and Rutledge*, 1990] place water pollution control expenditures in 1988 at $37.3 billion (1990 dollars), a level somewhat below our aggregate benefit estimate for 1990 of $46.7 billion. These expenditures are currently purchasing national water quality levels where most lakes, rivers, and streams are at least somewhere between fishable and swimmable and where a large number are swimmable. A small number of rivers and lakes, mostly near urban and industrial areas, are only of a boatable quality. *EPA* [1990] has estimated 1991 water pollution control expenditures will be $53.5 billion (1990 dollars) and has projected that such expenditures in the year 2000 will be $76.3 billion (1990 dollars). These markedly higher projected expenditures are due largely to moving from BPT (best practical technology) to BAT (best available technology) standards and the implementation of nonpoint source controls on agricultural and urban runoff. While a detailed discussion of the cost side of the Clean Water Act is beyond the scope of this paper, the interested readers should look at *Freeman* [1990] for a discussion of policy considerations, *EPA* [1990] for a detailed depiction of cost by industrial sector, and *Lake et al.* [1979] for a discussion of the distribution of cost by consumer groups.

It is important to note that our CV scenario did not address the possible impact of long-lived toxicants such as PCB's and heavy metals, which for the most part can no longer be legally discharged. Long-lived toxics primarily pose a problem, not because our respondents did not think that they were paying to prevent them but because if they failed to prevent them in any one year, the damages will carry over to other years. Our scenario poses a situation where estimates reflect benefits which may be curtailed by the cessation of expenditures or regained upon the resumption of such expenditures. Households may be willing to pay substantial amounts to control the release of such toxicants even if their willingness to pay for present water quality was zero if they had a desire for water quality in the future. *Freeman* [1990] notes that the earlier benefit studies also failed to account for this potentially important source of benefits.

CONCLUDING REMARKS

Measuring national water quality benefits using a national contingent valuation study offers several important advantages over alternative approaches. Among other advantages, it avoids the issue of how to aggregate benefits across disparate water-related activities and how to aggregate benefits over geographic areas. Furthermore, it easily incorporates nonuse benefits. We described how our study's key features were designed to address the problems of validity and reliability which might arise in trying to make valid and reliable inferences from data obtained from a national survey of the benefits of a national good.

Evidence supporting the validity of the data include: the favorable results of an experiment to test whether the respondents were unwittingly valuing a more general package of environmental improvements, the estimation of a parsimonious valuation function which was then used to successfully predict the results of a regional CV water benefits study, and comparisons with a variety of other studies which showed plausible relationships. While we believe the results of this study as a whole are sufficiently valid to be used for policy analysis, those who would do so should pay careful attention to how we defined the amenity and the context in which it would be provided.

Our estimates show that the potential annual benefits of swimmable quality water in the nation's freshwater lakes, rivers, and streams are large and in excess of the latest reported (1988) annual costs of the water quality improvement program. Looking to the future, however, total costs are projected to escalate well beyond total potential benefits owing to the higher marginal costs of bringing the remaining water bodies up to swimmable quality level. In revising the Clean Water Act, Congress can improve the future benefit-cost ratio by scaling back its goals, reducing the costs associated with obtaining the current goals, or both. Our work suggests that relaxing the uniform national swimmable quality goal, for example, by setting lower quality targets for certain stretches of large rivers with heavy industrial use would significantly reduce costs while resulting in only a small reduction in potential benefits. *Lyon and Farrow* [1993] discuss these issues in more detail. Finally, economists have long argued [e.g., *Hahn*, 1989] that it should be possible to reduce the cost across the board by moving from the grossly inefficient technology-based command and control approach adopted in 1972 to one of the economic incentive-based approaches. When Congress passed the Clean Water Act, almost any method of reducing projected effluent discharges would have increased the public's welfare. That is no longer true.

Acknowledgments. This research was supported in part by U.S. Environmental Protection Agency Cooperative Agreement CR 810224, to Resources for the Future. The author's alone are responsible for the views expressed in this paper.

REFERENCES

Boyle, K. J., R. C. Bishop, and M. P. Welsh, Starting points bias in contingent valuation surveys, *Land Econ.*, *61*, 188–194, 1985.
Bratton, D., and G. L. Rutledge, Pollution abatement and control expenditures, 1985–1988, *Surv. Curr. Bus.*, *70*, 32–38, 1990.
Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, Wadsworth, Belmont, Calif., 1984.
Cambridge Economics, Inc., *Contingent Valuation: A Critical Assessment, Symposium*, Cambridge, Mass., 1992.
Carson, R. T., N. J. Carson, A. Alberini, N. Flores, and J. Wright, *A Bibliography of Contingent Valuation Studies and Papers*, Natural Resource of Damage Assessment, Inc., La Jolla, Calif., 1993.
Cummings, R. G., D. S. Brookshire, and W. D. Schulze, *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*, Rowman and Littlefield, Totowa, N. J., 1986.
Dorfman, R., Incidence of the benefits and costs of environmental programs, *Am. Econ. Rev.*, *67*, 333–340, 1977.
Feenberg, D., and E. S. Mills, *Measuring the Benefits of Water Pollution Abatement*, Academic, San Diego, Calif., 1980.
Fisher, A. N., and R. Raucher, Intrinsic benefits of improved water quality: Conceptual and empirical perspectives, in *Advances in Applied Micro-Economics*, edited by V. K. Smith, JAI Press, Greenwich, Colo., 1984.
Freeman, A. M., *Air and Water Pollution Control: A Benefit Cost Assessment*, John Wiley, New York, 1982.
Freeman, A. M., Water pollution policy, in *Public Policies for Environmental Protection*, edited by P. Portney, Resources for the Future, Washington, D. C., 1990.
Goldberger, A. S., The interpretation and estimation of Cobb-Douglas functions, *Econometrica*, *35*, 464–472, 1968.

Goldfeld, S. M., and R. G. Quandt, *Nonlinear Methods in Econometrics*, North-Holland, New York, 1972.
Gramlich, F. W., The demand for clean water: The case of the Charles River, *Nat. Tax J.*, *30*, 183–194, 1977.
Greenley, D. A., R. G. Walsh, and R. A. Young, *Economic Benefits of Improved Water Quality: Public Perceptions of Option and Preservation Values*, Westview, Boulder, Colo., 1982.
Hahn, R. W., *A Primer on Environmental Policy Design*, Harwood, New York, 1989.
Hoehn, J. P., Valuing the multidimensional impacts of environmental policy: Theory and methods, *Am. J. Agric. Econ.*, *73*, 289–299, 1991.
Hoehn, J. P., and A. Randall. A Satisfactory benefit cost indicator from contingent valuation, *J. Environ. Econ. Manage.*, *14*, 226–247, 1987.
Hoehn, J. P., and A. Randall, Too many proposals pass the benefit cost test, *Am. Econ. Rev.*, *79*, 544–551, 1989.
Kahneman, D., *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*, edited by R. Cummings, D. Brookshire, and W. Schulze, Rowman and Littlefield, Totowa, N. J., 1986.
Kahneman, D., and J. L. Knetsch, Valuing public goods: The purchase of moral satisfaction, *J. Environ. Econ. Manage.*, *22*, 57–70, 1992.
Lake, E., W. M. Hanemann, and I. Strand, *Who Pays for Clean Water?: The Distribution of Water Pollution Control Cost*, Westview, Boulder, Colo., 1979.
Leone, R. A., and J. E. Jackson, The political economy of federal regulatory activity: The case of water-pollution controls, in *Studies in Public Regulation*, edited by G. Fromm, MIT Press, Cambridge, Mass., 1981.
Lyon, R. M., and S. Farrow, An analysis of clean water act issues, paper presented at the Association of Environmental and Resource Economists meeting, Anaheim, Calif., January 1993.
Madow, W. G., H. Nisselson, and I. Olkin, *Incomplete Data in Sample Surveys*, Academic, San Diego, Calif., 1983.
Mitchell, R. C., and R. T. Carson, An experiment in determining willingness to pay for national water quality improvements, report to the U.S. Environmental Protection Agency, Resources for the Future, Washington, D. C., 1981.
Mitchell, R. C., and R. T. Carson, A contingent valuation estimate of national freshwater benefits, report to the U.S. Environmental Protection Agency, Resour. for the Future, Washington, D. C., 1984.
Mitchell, R. C., and R. T. Carson, The use of contingent valuation data for benefit/cost analysis in water pollution control, report to the U.S. Environmental Protection Agency, Resour. for the Future, Washington, D. C., 1986.
Mitchell, R. C., and R. T. Carson, *Using Surveys to Value the Benefits for Public Goods*, Resources for the Future, Washington, D. C., 1989.
Randall, A., B. Ives, and C. Eastman, Bidding games for valuation of aesthetic environmental improvements, *J. Environ. Manage.*, *1*, 132–149, 1974.
Smith, V. K., Uncertainty, benefit-cost analysis, and the treatment of option value, *J. Environ. Econ. Manage.*, *14*, 283–292, 1987.
Smith, V. K., Arbitrary values, good causes, and premature verdicts, *J. Environ. Econ. Manage.*, *22*, 71–89, 1992.
Smith, V. K., and W. H. Desvousges, *Measuring Water Quality Benefits*, Kluwer, Academic, Hingham, Mass., 1986.
U.S. Environmental Protection Agency, Environmental investment: The cost of a clean environment, report to Congress, Washington, D. C., 1990.
White, H. L., A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, *48*, 817–838, 1980.
Willig, R. D., Consumer surplus without apology, *Am. Econ. Rev.*, *66*, 587–597, 1976.

Richard T. Carson, Department of Economics, D-008, University of California, San Diego, 210 Economics Building, La Jolla, CA 92093.

Robert Cameron Mitchell, Graduate School of Geography, Clark University, 950 Main Street, Worcester, MA 01610.

(Received December 10, 1991;
revised January 14, 1993;
accepted February 26, 1993.)