JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT 29, 238–251 (1995)

Assessing the Validity and Reliability of Contingent Values: A Comparison of On-Site Users, Off-Site Users, and Non-users[1]

JOHN C. WHITEHEAD

Department of Economics, East Carolina University, Greenville, North Carolina 27858

GLENN C. BLOMQUIST

Department of Economics, University of Kentucky, Lexington, Kentucky 40506

THOMAS J. HOBAN AND WILLIAM B. CLIFFORD

Department of Sociology, North Carolina State University, Raleigh, North Carolina 27695

Received January 28, 1994; revised June 16, 1994

Imperfect information in contingent markets can affect the validity and reliability of contingent values (willingness to pay, WTP) for changes in natural resource allocations. Survey respondents can acquire information about the natural resource to be valued from behavior such as on-site use of the resource, off-site use of the resource, and/or from the survey itself. In this paper we compare the validity and reliability of WTP from respondents who obtained information in these different ways. Using theoretical validity tests based on relative prices as well as income we find that WTP is valid for respondents with information from on-site and/or off-site use. Willingness to pay is less valid for respondents who had no information about the natural resource before they participated in the survey. Using an alternative form reliability test, we find that WTP is, in the technical sense, reliable for all respondents and somewhat more reliable for on-site users.

1. INTRODUCTION

The merit of using contingent values for policy analysis in large part is determined by their accuracy. Accuracy in contingent valuation depends on the relative validity and reliability of the valuation measure [23]. The validity of contingent

[1] The research on which this paper is based was financed in part by the United States Environmental Protection Agency and the North Carolina Department of Environment, Health, and Natural Resources, through the Albemarle-Pamlico Estuarine Study. Contents of this paper do not necessarily reflect the views and policies of the U.S. EPA or the N.C. DEHNR, nor does mention of trade names or commercial products constitute their endorsement by the U.S. or N.C. Government. The research reported in this paper was also funded (in part) by the North Carolina Agricultural Research Service and by an East Carolina University Faculty Senate Summer Research Grant.

This paper has benefited from the contributions of workshop participants at North Carolina State University, East Carolina University, the Agricultural University of Norway, and the University of Kentucky. We also acknowledge the contributions of two anonymous referees which have greatly improved this paper.

Address all correspondence to John Whitehead. Mathematical derivations, the survey instrument, further data description, and additional empirical results, which are referred to throughout this paper, are available upon request.

This paper was written while Glenn Blomquist was visiting at the Stockholm School of Economics.

0095-0696/95 $12.00

value, or stated willingness to pay (WTP), is the extent that the measured value corresponds to the theoretical definition of value. Technical reliability of stated WTP is the stability of the measure over time and the extent to which it is due to nonrandom sources. A contingent value measure can be valid but not reliable and vice versa.

There is considerable skepticism concerning the validity and reliability of contingent values, especially those values expressed by people who do not visit the natural resource [15]. The Cummings *et al.* [12] "reference operating conditions" suggest that non-use values are the type that are most likely to be inaccurate because respondents are unfamiliar with the resource. Based on these concerns we propose a differentiation among types of respondents (not among types of values) focusing on acquired information, or familiarity, categories. Willingess to pay statements can be made by

- on-site users: survey respondents who have been consumptive or nonconsumptive on-location users of the natural resource,
- off-site users: survey respondents who have read, seen a program, or talked about the natural resource, and
- non-users: survey respondents who have never seen or heard anything about the natural resource.

These categories are based on the premise that the lack of validity and/or reliability of WTP is due to imperfect information. The NOAA Panel on Contingent Valuation [1] focuses concern on WTP statements made by passive users, respondents who fit into both our off-site user and our non-user categories.

The primary purpose of this paper is to explore the contribution of acquired information in the assessment of the validity and reliability of contingent values. Another purpose is to introduce tests for validity assessment based on relative price effects and for reliability assessment based on an alternative form of the valuation question. These tests allow us to assess the comparative validity and reliability of WTP expressed by users, both on-site and off-site, and previously uninformed non-users of a natural resource.

2. INFORMATION AND WILLINGNESS TO PAY

Willingness to pay is the maximum amount of money respondents would give up in order to enjoy a natural resource change. For example, consider WTP for improvements in water quality and wildlife habitat in the Albemarle-Pamlico (A-P) estuarine system in North Carolina and a related natural resource, such as the Chesapeake Bay. A formal definition of willingness to pay is

$$\mathrm{WTP}_1 = e(p_1, p_2, q_1^0, q_2, u) - e(p_1, p_2, q_1', q_2, u), \tag{1}$$

where $e(\cdot)$ is the expenditure function, p_1 is the own-price, p_2 is the cross-price, q_1 is the quality of A-P system resources, q_2 is the quality of Chesapeake Bay resources, q_1^0 is the current (degraded) level of quality, q_1' is an improvement in quality, and u is utility. Expenditures necessary to maintain a specified utility level

decrease with an increase in quality (q^0 to q') so that WTP ≥ 0.[2] Whitehead and Blomquist [27] argue that information about the natural resource change is a necessary condition for positive WTP. Households gather information and then form values for natural resources which can then be expressed as

$$\text{WTP}_1 = e(p_1, p_2, q_1^0, q_2, u \mid I > 0) - e(p_1, p_2, q_1', q_2, u \mid I > 0), \qquad (2)$$

where I is information about the natural resource. When the household has information about the resource ($I > 0$) a positive WTP can be formulated (WTP $\geq$ 0). However, when the household has no information about the resource ($I = 0$), WTP is equal to zero

$$0 = e(p_1, p_2, 0, q_2, u \mid I = 0) - e(p_1, p_2, 0, q_2, u \mid I = 0). \qquad (2')$$

Positive WTP values result from observable behavior that leads to acquired information.[3]

Information about the resource can be acquired in at least three ways. First, a survey respondent who is an on-site user of the resource, perhaps a recreational angler or backpacker, will have information about the natural resource through personal experience. Second, a respondent may be an off-site user of the resource and acquire information through books, magazines, television, or conversations with friends. Third, a non-user will have no information about the natural resource before the survey but, during the process of participating in contingent valuation, acquires information from the survey instrument. Of course, on-site users can be off-site users also and both types of users could get new information via participation in a contingent market.

The validity and reliability of WTP depend on the amount and type of information that individuals have. In general, we suggest that the better the information, both acquired through personal experience and available in the survey instrument, the more valid and reliable are the WTP statements made by respondents. This proposition is similar to the reference operating condition of Cummings *et al.*, which specifies that the accuracy of WTP statements increases with the respondent's familiarity with the natural resource.

In related research, Boyle *et al.* [7] find that experience with white-water boating improves the accuracy of WTP for alternative river flows. Boyle *et al.* [8] ensure that most of their sample had no prior information about waterfowl in the Central Flyway so that most of the respondents were non-users. Using the number of birds as the measure of scope they find that non-use values for prevention of oil in ponds used for migration are not theoretically valid. Research is needed to assess the accuracy of WTP for resource changes with which respondents have no experience.

[2] Our model differs from that of Boyle and Bishop [6], which includes the prices of off-site use (which they call indirect use) goods such as magazines. It is not impossible that own-price and cross-price effects might be recovered from hedonic price equations for off-site use goods. We do not have this data and therefore suppress these terms in our model.

[3] In contrast, Bishop and Welsh [3] argue that values for obscure resources about which people have no information are theoretically possible. In the context of our model, their argument is that non-users can have positive and valid values for obscure resources if they could become informed. They further argue that these "conditional existence values" should be included in policy analysis because people are rationally ignorant due to the high cost of information.

3. VALIDITY AND RELIABILITY TESTS

The most common form of validity test in the contingent valuation literature is construct validity. Construct validity tests include convergent validity and theoretical validity tests. Convergent validity tests assess the convergence of the same contingent value construct measured in different ways. Numerous studies contribute to the evidence that contingent values stated by on-site users are convergent valid [10]. It is difficult to test the convergent validity of WTP stated by off-site users or non-users, however, since this type of behavior is difficult to observe.

Theoretical validity tests assess the relationship of the measure of contingent value and theoretical predictions. These tests are typically specified for the relationship between ability to pay and contingent values and can be conducted with all types of respondents [13]. Another test of theoretical validity, emphasized by the NOAA Panel [1], is the scope test where contingent values are tested to determine if they are monotonically increasing in quantity or quality [8, 11].

Theoretical definitions of contingent value, such as WTP, are usually modeled as differences in indirect utility or expenditure functions. These definitions should include relative prices as theoretical arguments of WTP. This paper represents an initial assessment of the theoretical validity of WTP with respect to relative price effects. Since the appropriate measure of price includes the value of time, we investigate the effects of changes in the own-price and a cross-price, with a positive value of time, on WTP for quality changes.

The most common form of technical reliability testing in contingent valuation is the test-retest method. The test-retest method includes a resurvey of respondents, presenting them with the same survey instrument, and compares responses in the same contingent market [18, 20, 24]. A limitation of test-retest reliability assessment is its high cost. A less costly test of reliability is the alternative or parallel form method [9, 25]. This test employs a single survey and two related valuation questions designed to be as similar as possible. For instance, if an open-ended WTP valuation question is initially presented, the alternative form question might be a closed-ended valuation question. Positive correlations between the measures estimated from the alternative forms are typically used as evidence of the reliability of both measures [19].

This paper reports results of an alternative form of reliability test in which the alternative form of the valuation question is not as closely related as open-ended and closed-ended forms of WTP. Specifically, elements of contingent valuation are purposely missing from our initial valuation question. This difference allows prediction of the direction of bias in the initial valuation. A contingent market is then constructed and a second valuation question, this time containing elements of a contingent market, is presented.[4] A positive and high correlation between the two

[4] Kealy and Turner [19] present results of what we would describe as an alternative form reliability test. However, they describe their test as one of convergent *validity*. There may also be some differences in interpretation of our reliability test as well. Since we can expect differences in the initial and subsequent valuations based on theory, our reliability test contains some elements of a theoretical validity test. We proceed in describing our test as one of reliability since the elements of a reliability test dominate [25].

responses, combined with consistent magnitudes of the two measures, would suggest that WTP is reliable.[5]

4. SPECIFICATION OF THEORETICAL VALIDITY TESTS

In order to derive theoretical predictions for validity tests, consider equation (2) assuming $I > 0$ for respondents and suppressing I for simplicity. Assume the reference level of utility is $u = v(p_1, p_2, q_1', q_2, y)$, where y is income and $v(\cdot)$ is the indirect utility function. Substitution of the indirect utility function into Eq. (2) yields

$$\text{WTP}_1 = e\left[p_1, p_2, q_1^0, q_2, v(p_1, p_2, q_1', q_2, y)\right] - y, \tag{3}$$

which is called the variation function by McConnell [21].

Comparative static analysis of the variation function can be used to specify validity tests.[6] For example, WTP for quality changes is increasing in income if quality is a normal good,

$$\frac{\partial \text{WTP}_1}{\partial y} = \gamma - 1 > 0, \tag{4}$$

where

$$\gamma = \frac{\partial e(q_1^0)/\partial v}{\partial e(q_1')/\partial v} > 1$$

if q is a normal good (which we will assume throughout the remainder of the paper).

Other relevant arguments of the variation function are the own price and cross price. The effect of the own price on WTP is

$$\frac{\partial \text{WTP}_1}{\partial p_1} = x_1^m(p_1, p_2, q_1^0, q_2, y) - \gamma\left[x_1^m(p_1, p_2, q_1', q_2, y)\right] < 0, \tag{5}$$

where $x^m(\cdot)$ is the uncompensated (Marshallian) demand function for recreation trips. WTP is inversely related to the own price of on-site use because recreation demand, $x_1^m(\cdot)$, is positively related to quality if trips and quality are complements.

The relationship between the own-price effect and the change in recreation demand can be seen by recognizing that $\Delta x_1 = x_1^m(p_1, p_2, q_1', q_2, y) - x_1^m(p_1, p_2, q_1^0, q_2, y)$, where Δx_1 is the change in the number of recreation trips. Solving (5) for $x_1^m(p_1, p_2, q_1^0, q_2, y)$ and substituting into Δx_1 yields $\Delta x_1 = [(1 -$

[5] A criticism of the test-retest method is that the correlation may be upwardly biased if respondents, in the retest questionnaire, anchor their subsequent WTP responses to their initial WTP responses. Correlations of open-ended and closed-ended measures of WTP can be upwardly biased if respondents anchor their answers to the second valuation question to their first answer. Of course, correlations of responses to the alternate forms of valuation questions may also be upwardly biased if respondents anchor their subsequent valuation to their initial valuation.

[6] For the sake of brevity we skip several intermediate steps in this section (see Whitehead [26]). The income result follows McConnell [21].

$\gamma) \cdot x_1^m(p_1, p_2, q_1', q_2, y) - \partial \mathrm{WTP}_1/\partial p_1]$. So, $\Delta x_1 \to -\partial \mathrm{WTP}_1/\partial p_1$ as $\gamma \to 1$. The own-price result (5) also holds for those off-site users or non-users who would participate in recreation after a quality improvement. For off-site users and non-users the quality improvement would increase recreation demand so that the currently faced own price is no longer equal to the choke price.[7]

For off-site users or non-users who face the choke price at both the original and improved quality levels,

$$\frac{\partial \mathrm{WTP}_1}{\partial p_1} = 0. \tag{6}$$

The own-price effect will equal zero if the improved quality does not induce recreation participation.

The cross-price effect is

$$\frac{\partial \mathrm{WTP}_1}{\partial p_2} = x_2^m(p_1, p_2, q_1^0, q_2, y) - \gamma[x_2^m(p_1, p_2, q_1', q_2, y)] \gtreqless 0. \tag{7}$$

If x_1 and x_2 are substitutes and $\gamma \to 1$ then the cross-price effect is positive since increases in q_1 will increase x_1 and decrease x_2. If x_1 and x_2 are complements, then the cross-price effect is negative since increases in q_1 will increase x_1 and x_2. If the resources are unrelated in consumption then the cross-price effect is negative because $\gamma > 1$; however, this effect will be small as $\gamma \to 1$. If the number of trips to the related resource is equal to zero at either quality level, the cross-price effect is equal to zero.

The relationship between the cross-price effect and the change in recreation demand at the related site can be found by recognizing that $\Delta x_2 = x_2^m(p_1, p_2, q_1', q_2, y) - x_2^m(p_1, p_2, q_1^0, q_2, y)$. Solving (7) for $x_2^m(p_1, p_2, q_1^0, q_2, y)$ and substituting into Δx_2 yields $\Delta x_2 = [(1 - \gamma) \cdot x_2^m(p_1, p_2, q_1', q_2, y) - \partial \mathrm{WTP}_1/\partial p_2]$. As $\gamma \to 1$, $\Delta x_2 \to -\partial \mathrm{WTP}_1/\partial p_2$.

5. CONTINGENT VALUATION AND THE RELIABILITY TEST

Our alternative form reliability test compares responses to a WTP question and a question that elicits the "worth" of the resource. The alternative form test was conducted in the context of a study about improvements in water quality and wildlife habitat in the A-P estuarine system [17]. We designed the survey to first describe the A-P estuarine system and then elicit information about knowledge and on-site recreational use of the system. On-site users of the A-P system are those respondents who stated that they had engaged in fishing, swimming, boating, or some other activity during the past year on the Albemarle Sound, the Pamlico Sound, or the feeder rivers. Off-site users are those respondents who answered "a little," "some," or "a lot" to the question "how much have you heard or read about the resources, uses, and problems of the A-P system?" and who are not on-site

[7] For off-site users and non-users, since the choke price implicitly depends on quality, $\partial p_1^c/\partial q_1 > 0$, where p_1^c is the choke price for recreation, $x_1^m(\cdot, q_1^0) = 0$ and $x_1^m(\cdot, q_1') \geq 0$. As quality improves, demand may increase and off-site users and non-users may enter the recreation market.

users. Non-users are those who answered "nothing" to the off-site use question above and who are not on-site users.

Through a series of questions, the different individual uses of the system such as fishing, boating, and swimming are then described. Other collective uses of the system such as water supplies and commercial fishing, as well as knowing that others including future generations may use and enjoy the A-P system are also described. Next, the alternative form of the valuation question elicits the WORTH of the A-P system:

> Now I would like to know how much you and your household value water quality and fish and wildlife habitat in the A-P system. This system provides a number of benefits, such as recreation, tourism, commercial fishing, food production, and wildlife habitat. Many people also like to know the rivers and sounds will be protected for future generations. Would it be *worth* (emphasis added) $A each year to you and your household to make sure water pollution does not get worse and wildlife habitat remains the same in the A-P system?

The starting point ($A) took on twelve values with a random start ranging from $5 to $100 (5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, and 100). The range of starting points was pretested using an iterative bidding procedure [16]. Once initial yes or no answers were given, respondents were asked follow-up questions to further elicit a range of WORTH. An open-ended question, "What is the most that protecting this system would be worth to you and your household each year?", was presented to respondents who did not agree with any of the dollar amounts. The WORTH question does not imply a payment obligation or payment vehicle.

A series of questions concerning the ways that the A-P system can become polluted and the type of government programs that can improve water quality is presented next. These questions were designed to describe the management program and to make the payment obligation explicit. Next, our WTP question was in the form of an iterative political market which presents survey respondents with a referendum vote[8]:

> We already pay for the types of government programs we've just discussed through federal, state, and local taxes. However, government will need more money if the water quality and fish and wildlife habitat in the A-P system are to be protected. This money would pay for state and local programs to control pollution, monitor water quality, protect habitat, and educate people. The goal would be to make sure water pollution does not get worse and habitat remains the same. Would you and your household be *willing to pay* (emphasis added) $A, each year, in *higher taxes* (emphasis added), for these programs, if you knew the money would be used to protect the A-P system?

We followed a similar iterative bidding procedure as in the WORTH section with identical starting-point tax payments and iteration procedures. The open-ended question "What is the most that you and your household would be willing to pay each year...?" is presented to respondents who did not agree to pay any of the dollar amounts. The WTP question makes explicit the payment obligation and payment vehicle.

[8] The A-P survey creates a contingent market for a government program designed to improve water quality and wildlife habitat assuming that water quality and wildlife habitat will decrease from the current level. In a preliminary study, Hoban and Clifford [16] found empirical support for this assumption.

Since the payment vehicle and obligation are not explicitly mentioned, WORTH is not necessarily income constrained and is more hypothetical. Therefore WORTH should be greater than or equal to WTP. A weak test of the alternative form reliability of WTP is the correlation between WORTH and WTP. Positive and significant correlations between WORTH and WTP lend some reliability to the WTP statement. A stronger test of reliability is the correlation plus a difference in means test. A mean WORTH significantly greater than mean WTP lends reliability to the WTP statement. Another stronger test of reliability is the chi-square test for count data of the proportions of respondents who state WTP less than or equal to WORTH.

6. THE DATA

Our survey design was cross-sectional, utilizing a random sample of households with telephones. The universe for the survey was defined as the 100 counties in North Carolina and the 16 counties/independent cities in Virginia within the watershed of the A-P system. The 1990 population of the region was 7,442,684 individuals and, based on the average number of persons per household, 2.54, the number of households was 2,930,190. The counties making up the universe were stratified into five standard geographic regions in order to sample the entire region affected by the estuarine system and over-sample the sparsely populated, North Carolina coastal plain. Phone numbers for the sample were selected using a random digit dialing technique. Telephone interviews were completed with 1133 households, constituting an overall response rate of 70.5%.

In Table I we present the means of the variables by user status. The WTP data we analyze do not include protest or outlier responses.[9] The WTP data are mixed continuous/interval with 13% zero WTP responses, 63% interval data responses, and 24% open-ended responses greater than $100. For WTP values less than $100 and greater than $5 the mid-point of each $5–$10 interval is used as an estimate of the WTP within the interval. WTP values equal to zero and greater than $100 are included as point estimates.

Construction of the own-price and cross-price variables is consistent with the recreation demand literature. While we do not have any direct evidence that respondents use the Chesapeake Bay, we chose the Chesapeake Bay as the potentially related natural resource due to its proximity and similarity of characteristics and services provided.[10] The round-trip distance to both the A-P system and the Chesapeake Bay is calculated as the driving distance from the population center of the respondent's county to the nearest water access. The opportunity cost of time is equal to the wage rate and the average driving speed is the North Carolina state average.

[9] For zero dollar respondents the follow-up question "Why would you not be willing to pay anything?" is presented. Respondents who reject the contingent market institution, existing governmental institutions, or the payment vehicle are considered protests. Outliers are those respondents who could not formulate a WTP answer and whose WTP was coded as "greater than $995 or priceless." A WTP greater than $995 is $300 more than the next highest WTP response ($700) and at least 3% of average household income for our sample. Fifty-six protest responses and 29 outlier responses are deleted. There are also 15 item non-responses to the WTP question.

[10] Indirect and weak evidence that respondents use the Chesapeake Bay is that respondent use of "other water bodies" for outdoor recreation and the price of a Chesapeake Bay trip is negatively related and statistically significant in a recreation demand model.

TABLE I
Data Summary[a]

Variable	On-site users	Off-site users	Non-users
WTP	$71.13	$55.76	$49.46
	(78.92)[b]	(59.54)	(47.19)
WORTH[c]	$77.37	$69.45	$52.88
	(76.80)	(81.10)	(54.58)
Own-price[d]	$121.42	$142.38	$156.67
	(106.08)	(116.03)	(146.09)
Cross-price[e]	$143.19	$169.87	$173.04
	(129.78)	(144.60)	(175.82)
Income[f]	$40,325.36	$32,944.91	$31,659.01
	(27,946.53)	(25,772.96)	(29,803.74)
A	$46.15	$47.52	$46.40
	(31.40)	(31.24)	(30.59)
n	308	518	207

[a] 1990 dollars.
[b] Standard deviation in parentheses.
[c] n = 262, 454, 192 for on-site users, off-site users, and non-users due to item non-response.
[d] The dollar and time costs of a trip to the Albemarle-Pamlico estuarine system.
[e] The dollar and time costs of a trip to the Chesapeake Bay.
[f] Total income before taxes including wages and salaries, pension dividends, net farm income, and government payments.

Tests for differences in means of the characteristics of respondents are conducted which suggest that off-site users and non-users are similar but that on-site users are different from both groups. For on-site users, each difference in mean is significantly different from zero when compared to both groups except for the difference in WORTH for the on-site users and off-site users. The only statistically significant difference in means between the off-site users and the non-users is for the WORTH variable. The tax amount ($\$A$) variable is insignificantly different across groups.

7. EMPIRICAL RESULTS

The validity of WTP is assessed through Tobit regression analysis (Table II).[11] Reliability is assessed through correlation, differences in means, and proportions

[11] Data are weighted in order to adjust for the regional stratification of the sample. The choice of estimation technique is among the grouped data, Tobit, and Logit or Probit (based on the responses to the initial dichotomous choice question) regression. The initial dichotomous choice response data, however, suffer from the "fat tails" problem. Although the starting points were pretested in Hoban and Clifford [16], 59% of the sub-sample presented with the highest tax amount ($100) agreed to pay. We present the Tobit regressions due to the tightness of our WTP intervals and the censoring of the data [29]. The Tobit coefficients presented in Table II are adjusted so that they have a direct economic interpretation [22]. The adjusted coefficients (β^A) are $\beta^A = \sigma \cdot F(z) \cdot \beta^T$, where σ is the standard error of the regression, $F(z)$ is the unit cumulative normal distribution, and β^T are the Tobit coefficients.

Starting point bias results when respondents consider the initial dollar amount offered as an implied "correct" value [5]. Since the starting points were randomly assigned, the correlations between starting points and other independent variables are low. Starting point bias does not affect our validity tests.

TABLE II
Validity Tests: Tobit Coefficients, Dependent Variable = WTP

Variable	On-site users	Off-site users	Non-users
Intercept	−1.54	−5.43	11.49*
	(−0.17)[a]	(−1.11)	(1.92)
Own price	−0.19**	−0.13**	−0.0028
	(−2.34)	(−2.48)	(−0.034)
Cross price	0.066	0.15***	0.040
	(1.13)	(3.10)	(0.68)
Income	0.00102***	0.00051***	−0.000054
	(5.15)	(4.10)	(−0.30)
A	0.65***	0.49***	0.42***
	(5.35)	(7.41)	(4.34)
σ[b]	80.63***	56.00***	50.87***
	(23.61)	(28.81)	(17.70)
Log likelihood function	−1684.60	−2517.20	−967.12
n	308	518	207

Note. ***, **, and * indicate significance at the p = .01, .05, and .10 levels.
[a] t statistic in parentheses.
[b] Standard error of the regression.

tests (Table III). A comparative analysis of respondents who obtained information from on-site use, off-site use, and only from the survey itself (the non-users) is performed.

For both the on-site user and off-site user samples, the own-price coefficient is the correct sign and significantly different from zero.[12] WTP decreases by $.19 for on-site users and decreases by $.13 for off-site users when own price increases by $1. The own-price coefficient for the non-user sample is negative but insignificantly different from zero. These results are consistent with predictions from theoretical validity tests. The sizes of the own-price coefficients are plausible. With a quality

[12] This result does not hold for the on-site user sample when protest and outlier responses are included in the WTP regression because the own-price coefficient becomes insignificant. All other validity and reliability tests are stable whether the protests or outliers are included or not.

TABLE III
Reliabiltiy Tests

Test[a]	On-site users	Off-site users	Non-users
r(WORTH, WTP)	0.92	0.64	0.70
(WORTH − WTP)[b]	$10.68	$14.55	$6.26
	(5.59)[c]	(7.25)	(2.11)
%(WORTH ≥ WTP)[d]	76.95	78.38	75.85
	(9.46)[e]	(12.92)	(7.44)
n	262	454	192

[a] All results are statistically significant at the p = .01 level.
[b] Wilcoxon signed-rank test.
[c] z statistic in parentheses.
[d] Test of proportions based on the binomial distribution for count data.
[e] Chi-square statistic in parentheses.

improvement, the typical on-site user would take .19 additional recreation trips per year with a quality improvement and the typical off-site user would take .13 recreation trips to the A-P system per year.[13] The typical non-user would not participate in recreation with improved quality.

Chesapeake Bay recreation trips are substitutes for A-P system recreation trips for off-site users since the cross-price coefficient is positive. The typical off-site user would be willing to pay $.15 more for a quality improvement with a $1 increase in the price of a recreation trip to the Chesapeake Bay. The size of the cross-price coefficient is plausible. The typical off-site user would make .15 fewer recreation trips to the Chesapeake Bay with the quality improvement in the A-P system. The cross-price coefficients for the on-site user and non-user samples are insignificantly different from zero, suggesting on-site use of the Chesapeake Bay would not change. These results are also consistent with predictions from theoretical validity tests.

The income coefficient is positive and significant in both on-site and off-site user models, suggesting that households that have information about the resource consider their income constraints when stating WTP. Income elasticities, calculated using the means of income and WTP, for on-site users (.58) and off-site users (.30) are both plausible. The income coefficient for the non-user sample is negative and insignificantly different from zero, suggesting that households that have no information about the resource before the survey do not consider their income constraints when stating WTP.

Likelihood ratio tests for equivalence of vectors of coefficients are next conducted for pairs of user groups. Each coefficient vector is significantly different from the others at the $p = .01$ level, which suggests that respondents in each group formulate their WTP response in different ways. Since means of WTP, own-price, cross-price, and income for off-site users and non-users are not significantly different, theoretical validity tests are essential in order to distinguish between the two groups. The difference in mean WTP between the on-site users and off-site users is due to a combination of differences in variables and differences in coefficient vectors.

For the reliability tests we analyze cases that report both WORTH and WTP values. The correlation coefficient (r) between WORTH and WTP is high and statistically different from zero for each of the user groups. The correlation is highest for on-site users. A stronger reliability test is a comparison of differences in means between WORTH and WTP combined with the correlation coefficient. This test measures the consistency of responses and the magnitude of the responses. In all three samples, mean WTP is less than mean WORTH according to the Wilcoxon signed-rank test statistic [14]. For each group, differences in means between WORTH and WTP are significantly different from zero. Another stronger reliability statistic is the proportions test in which the proportions of reliable WTP

[13] Recall that the own-price effect measures (the negative of) the quality-induced recreation demand shift and that this is true when $\gamma \to 1$. An estimate of γ can be found from the coefficient on the income variable which is equal to $\gamma - 1$. For the on-site user and off-site user samples, $\gamma = 1.00102$ and 1.00051, respectively. The off-site user result suggests that respondents are initially (with low quality) priced out of the recreation market but with an improvement in quality the recreation demand curve shifts right to where the own-price is less than the choke price.

responses are tested against a random distribution. The chi-square statistics indicate that the WTP responses for each of the user groups is reliable.[14]

Comparison of reliability between the three user groups finds some evidence that on-site user WTP responses are more reliable than the others. A test for differences in correlations reveals that the correlation for the on-site users is statistically significantly greater than that for the off-site users and non-users [4]. There is no statistically significant difference in the correlations between the off-site users and non-users. However, using a chi-square test for differences in proportions, no statistical difference is found for the count data [14]. We have no theory to suggest the appropriate difference in means between WTP and WORTH and do not pursue this comparison.

8. CONCLUSIONS

In this paper we introduce tests for the validity and reliability of WTP. We find that relative prices and income have predictable effects on WTP for both on-site and off-site users of the resource. The same cannot be said for non-users who had no information about A-P system resources prior to the survey. Considering especially the test for income effects, we find that the non-user group does not state WTP in theoretically predicted ways. We find that WTP stated by on-site users, off-site users, and non-users is reliable based on an alternative form reliability test. One comparative reliability test statistic, based on correlations, suggests that WTP stated by on-site users is more reliable than that stated by off-site users and non-users.

We find that the information acquired by respondents before the survey information is presented can be a key determinant of validity and reliability. In general, validity and reliability increase with familiarity. These results imply that contingent markets presented to respondents who are unfamiliar with the resource will generate less accurate statements of WTP. When it is suspected that potential respondents are unfamiliar with the natural resource, greater effort could be made to inform respondents about characteristics of the choice problem, such as the quality of the resource and related environmental goods [2, 28]. In this context the divergence of WTP caused by information treatment groups can be interpreted as a sign of increasing validity of WTP.

Our results suggest that the set of usable WTP responses includes not only on-site users of a resource but also off-site users, those respondents who have expended their own resources and acquired information about the resource prior to being presented a contingent market. This is only part of the group of passive users referred to by the NOAA Panel on Contingent Valuation [1]. The greatest

[14] We also conduct regression analysis to determine whether the differences between WTP and WORTH are for the expected reasons or due to two intervening variables. The two variables have no effect on the correlation between WTP and WORTH. Another concern is if WORTH and WTP are different concepts to the average person. We test this by regressing WORTH on the theoretical determinants of WTP found in our validity test. These results strongly suggest that the concept of WORTH is less of an economic concept than WTP, especially for off-site users.

challenge is contingent valuation by people with no prior information about the natural resource.

REFERENCES

1. K. Arrow, R. Solow, E. Leamer, P. Portney, R. Radner, and H. Schuman, Report of the NOAA Panel on Contingent Valuation, *Fed. Register* **58**, 4602–4614 (1993).
2. J. C. Bergstrom, J. R. Stoll, and A. Randall, The impact of information on environmental commodity valuation decisions, *Amer. J. Agr. Econom.* **72**, 614–621 (1990).
3. R. C. Bishop and M. P. Welsh, Existence values in benefit-cost analysis and damage assessment, *Land Econom.* **68**, 405–417 (1992).
4. H. M. Blalock, Jr., "Social Statistics," McGraw-Hill, New York (1972).
5. K. J. Boyle, R. C. Bishop, and M. P. Welsh, Starting point bias in contingent valuation bidding games, *Land Econom.* **61**, 188–194 (1985).
6. K. J. Boyle and R. C. Bishop, Valuing wildlife in benefit-cost analyses: A case study involving endangered species, *Water Resour. Res.* **23**, 943–950 (1987).
7. K. J. Boyle, M. P. Welsh, and R. C. Bishop, The role of question order and respondent experience in contingent-valuation studies, *J. Environ. Econom. Management* **25**, S80–S99 (1993).
8. K. J. Boyle, W. H. Desvousges, F. R. Johnson, R. W. Dunford, and S. P. Hudson, An investigation of part-whole biases in contingent-valuation studies, *J. Environ. Econom. Management* **27**, 64–83 (1994).
9. E. G. Carmines and R. A. Zeller, "Reliability and Validity Assessment," Sage, Beverly Hills, CA (1980).
10. R. T. Carson, N. E. Flores, K. M. Martin, and J. L. Wright, "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods," Discussion Paper 94-07, University of California, San Diego, Department of Economics (1994).
11. R. T. Carson and R. C. Mitchell, Sequencing and nesting in contingent valuation surveys, *J. Environ. Econom. Management* **28**, 155–173 (1995).
12. R. G. Cummings, D. S. Brookshire, and W. D. Schulze, "Valuing Environmental Goods," Rowman and Allenheld, Totowa, NJ (1986).
13. W. H. Desvousges, V. K. Smith, and A. Fisher, Option price estimates for water quality improvements: A contingent valuation study for the Monongahela River, *J. Environ. Econom. Management* **14**, 248–267 (1987).
14. J. E. Freund and R. E. Walpole, "Mathematical Statistics," 3rd ed., Prentice-Hall, Englewood Cliffs, NJ (1980).
15. J. A. Hausman, Ed., "Contingent Valuation: A Critical Assessment," North-Holland, New York (1993).
16. T. Hoban and W. Clifford, "Public Attitudes Toward Water Quality and Management Alternatives in the Albemarle-Pamlico Estuarine System," Phase I Report, Project No. 90-27, Albemarle-Pamlico Estuarine Study (1991).
17. T. Hoban, and W. Clifford, "Public Attitudes Toward Water Quality and Management Alternatives in the Albemarle-Pamlico Estuarine System," Phase II Report, Project No. 92-13, Albemarle-Pamlico Estuarine Study (1992).
18. M. J. Kealy, M. Montgomery, and J. F. Dovido, Reliability and predictive validity of contingent values: Does the nature of the good matter? *J. Environ. Econom. Management* **19**, 244–263 (1990).
19. M. J. Kealy and R. W. Turner, A test of the equality of close-ended and open-ended contingent valuations, *Amer. J. Agr. Econom.* **75**, 321–331 (1993).
20. J. B. Loomis, Comparative reliability of the dichotomous choice and open-ended contingent valuation techniques, *J. Environ. Econom. Management* **18**, 78–85 (1990).
21. K. E. McConnell, Models for referendum data: The structure of discrete choice models for contingent valuation, *J. Environ. Econom. Management* **18**, 19–34 (1990).
22. J. F. McDonald and R. A. Moffitt, The uses of Tobit analysis, *Rev. Econom. and Statist.* **61**, 318–321 (1979).
23. R. C. Mitchell and R. T. Carson, "Using Surveys to Value Public Goods: The Contingent Valuation Method," Resources for the Future, Washington, DC (1989).

24. S. D. Reiling, K. J. Boyle, M. L. Phillips, and M. W. Anderson, Temporal reliability of contingent values, *Land Econom.* **66**, 128–134 (1990).
25. R. Singleton, Jr., B. C. Straits, M. M. Straits, and R. J. McAllister, "Approaches to Social Research," Oxford, New York (1988).
26. J. C. Whitehead, Willingness to pay for quality improvements: Comparative statistics and interpretation of contingent valuation results, *Land Econom.* **71**, 207–215 (1995).
27. J. C. Whitehead and G. C. Blomquist, A link between behavior, information, and existence values, *Leisure Sci.* **13**, 97–110 (1991).
28. J. C. Whitehead and G. C. Blomquist, Measuring contingent values for wetlands: Effects of information about related environmental goods, *Water Resour. Res.* **27**, 2523–2531 (1991).
29. J. C. Whitehead, T. J. Hoban, and W. B. Clifford, Measurement issues with iterated continuous/interval contingent valuation data, *J. Environ. Management* **43**, 129–139 (1995).