NCEE
NATIONAL CENTER FOR
ENVIRONMENTAL ECONOMICS

A Stated Preference Study of the Chesapeake Bay and Watershed Lakes

Chris Moore, Dennis Guignet, Kelly Maguire, Chris Dockins,
and Nathalie Simon

Working Paper Series

Working Paper # 15-06
November, 2015

U.S. Environmental Protection Agency
National Center for Environmental Economics
1200 Pennsylvania Avenue, NW (MC 1809)
Washington, DC 20460
http://www.epa.gov/economics

A Stated Preference Study of the Chesapeake Bay and Watershed Lakes

Chris Moore, Dennis Guignet, Kelly Maguire, Chris Dockins, and
Nathalie Simon

NCEE Working Paper Series
Working Paper # 15-06
November, 2015

A Stated Preference Study of the Chesapeake Bay and Watershed Lakes*

Chris Moore†‡, Dennis Guignet†, Kelly Maguire†, Chris Dockins†,

and Nathalie Simon†

Abstract

The Chesapeake Bay is the largest estuary in the United States and the third largest in the world. The surrounding Watershed encompasses 64,000 square miles, and is home to about 18 million people. There have been numerous studies measuring the value of different components of the Chesapeake Bay but no study or set of studies provides a comprehensive estimate of the values associated with the improvements likely to result from recently implemented pollution limits. To fill this gap we developed a stated preference (SP) survey that uses a discrete choice experiment response format to examine households' willingness to pay (WTP) for water quality improvements in the Chesapeake Bay. During extensive focus group testing, care was taken to identify the environmental attributes that are most important to both users and non-users and to quantitatively describe these attributes using understandable and tangible metrics. The survey was mailed to a stratified random sample of households across 17 states in the eastern U.S. and the District of Columbia. This paper reports the results of the empirical analysis, including marginal WTP estimates for each environmental attribute and total WTP for the expected outcome of the pollution reduction program. A comparison of WTP across households suggests that a substantial portion of the benefits can be attributed to nonusers. Results also show that benefits from improving water quality in freshwater lakes in the Watershed are an important ancillary benefit likely to result from reducing pollution in the Bay.

JEL Classification: Q51, Q53
Keywords: Chesapeake Bay, choice experiment, stated preference, TMDL, water quality

*We thank Elena Besedin, Maureen Cropper, and Alan Krupnick for valuable guidance and feedback throughout the development of the survey used in this study. We also thank Kevin Boyle, John Whitehead, and Robert Johnston for comments on earlier drafts of the survey instrument, and thank Julie Hewitt, George Parsons, Dan Phaneuf, and Greg Poe for comments on earlier paper drafts. Finally, we thank Bryan Milstead, Gary Shenk, and Steve Newbold for providing the hydrological and ecological modeling results we used to forecast future baseline and policy conditions when estimating economic benefits.

†US EPA, National Center for Environmental Economics. The views expressed in this article are those of the authors and do not necessarily represent those of the US EPA. Although research in this paper may have been funded entirely or in part by the US EPA, it has not been subjected to formal Agency peer and policy review. No official Agency endorsement should be inferred.

‡ Corresponding author. 1200 Pennsylvania Ave, NW (1809T), Washington, DC 20460, 202-566-2348, moore.chris@epa.gov

1. INTRODUCTION

The Chesapeake Bay is the largest estuary in the United States and the third largest in the world. The surrounding Watershed encompasses 64,000 square miles in parts of six states and the District of Columbia and is home to about 18 million people. The Chesapeake Bay's unique set of ecological and cultural elements has motivated efforts to preserve and restore its condition for more than 25 years. In December 2010 a pollution budget that puts limits on nutrient and sediment inflows from the surrounding Watershed, called a Total Maximum Daily Load (TMDL), was established for compliance with the Clean Water Act. The Chesapeake Bay TMDL requires loadings of nitrogen, phosphorus, and sediment to be reduced by 25%, 24%, and 20%, respectively[1].

Benefits from meeting the TMDL will accrue to (i) those who live near the Bay or visit for recreation; (ii) those who live near or visit lakes and rivers in the Watershed; and (iii) those who live further away and/or may never visit, but have a concern for these water bodies. While benefits from the first two categories can be measured using revealed preference approaches, only stated preference methods can capture non-use values, which may be substantial.

We conduct a stated preference survey of residents living in 17 east coast states and the District of Columbia. The survey employed a discrete choice experiment (DCE) response format to estimate willingness to pay (WTP) for improvements in five environmental attributes: water clarity in the Bay; populations of three iconic Chesapeake species - striped bass, blue crab, and

[1] Relative to 2009 levels.

eastern oysters; and the condition of freshwater lakes in the Chesapeake Bay Watershed. The DCE response format and our experimental design allow us to estimate marginal WTP for each attribute, and to estimate total WTP for outcomes expected from the TMDL.

The remainder of the paper is organized as follows. Section 2 provides background information and Section 3 describes the survey design, including the survey instrument and attributes, and the survey development process. Section 4 details the survey implementation and summarizes the data. Analysis of the DCE data including marginal WTP estimation is described in section 5. Household and total WTP results are presented in section 6. Section 7 presents the results of the sensitivity analyses and section 8 concludes.

2. BACKGROUND

There are many US-based studies in the environmental economics literature that quantify benefits or estimate willingness to pay (WTP) associated with various types of water quality and aquatic ecosystem changes.[2] For reviews of the literature see Johnston et al. (2005) and Van Houtven et al. (2007). However, few studies provide the necessary information to quantify values associated with clean-up programs in the Chesapeake Bay. Johnston et al. (2002) contains a coordinated set of four studies designed to estimate benefits of resource preservation and restoration decisions in the Peconic Estuary system in New York. One of these studies uses a choice experiment to estimate household WTP values for an additional acre of farmland, undeveloped land, wetlands, shellfishing areas, and aquatic grasses. It is difficult to directly compare the values from this study

[2] We also acknowledge important, related work done in Australia (e.g., Kragt and Bennett 2011, Windle and Rolf 2004) and Europe (e.g., Brouwer 2008, Metcalf et al. 2012).

to our own because Johnston et al. express results in terms of annual value per acre per household, but estimates for wetlands, shellfish, and eelgrass range from a low of $0.04 (shellfish areas) to $0.17 (eelgrass) depending upon model specification.[3] Nor would it be appropriate to transfer these values to the Chesapeake Bay because the resources, affected populations, baseline condition and policy outcomes are too dissimilar.

Cropper and Isaac (2011) review valuation studies relevant to improving water quality in the Chesapeake Bay and find very few attempts to estimate total or non-use benefits of improved water quality in the Chesapeake Bay. Bockstael et al. (1988, 1989) estimate willingness to pay to make the Bay “swimmable” for those respondents who considered that it was not acceptable for swimming. Lipton (2004) estimates willingness to pay of non-users for restoring oyster reefs in the Chesapeake Bay using a broad, but non-random, sample that covered most Mid-Atlantic States. Hicks et al. (2008) examines a broader variety of environmental outcomes associated with reduced sediment and nutrient loads in the Bay, however cost is not included as an attribute in the survey, so no WTP estimates could be inferred. Together, these studies provide limited information about values for new clean-up programs in the Bay.

Practices put in place to reduce nutrient and sediment pollution in coastal estuaries will generally reduce pollution in freshwater lakes and reservoirs in the same watershed (Moore et al. 2011). Banzhaf et al. (2006) use contingent valuation (CV) to estimate willingness to pay to reduce acidification at a collection of lakes in the Adirondacks. A referendum style CV question was also

[3] Updated to $2013.

used by Herriges et al. (2010) to estimate WTP for water quality improvements at a single lake, characterized by water clarity, color, odor, health concerns, and variety and quantity of fish. Roberts et al. (2008) employ a DCE for a single lake using algae bloom status and water level as attributes presented both probabilistically and with certainty. Viscusi et al. (2008) include a DCE for freshwater water quality improvements in lakes and rivers in a nationally representative sample. The attribute considered by respondents is the percentage of lake acres and river miles with water quality safe for all uses (i.e., with "good" water quality). Finally, Phaneuf et al. (2013) employ both a CV referendum question and a DCE approach to value improvements in lake water quality within respondents' home state. Water quality was characterized by five trophic categories, described in terms of water color, clarity, odor, type and abundance of fish, and size and frequency of algae blooms. The CV results from Phaneuf et al. are further developed in Van Houtven et al. (2014). In many ways this is the most similar to our own study, with a focus on broad programs to improve water quality in lakes across a region.

Our study fills a key gap in the literature by valuing a broad set of water quality outcomes for the Chesapeake Bay Watershed, including the Chesapeake Bay itself, the tidal reaches of its tributaries, and freshwater lakes in the Watershed. Such an approach is essential to evaluating the benefits of policies such as the TMDL, which have broad effects across all of these water bodies. We are also able to obtain distinct values for water quality improvements for both users and non-users of the Chesapeake Bay and Watershed lakes.

3. SURVEY DESIGN

The survey instrument was designed through extensive focus groups held in several locations, both within and outside of the Chesapeake Bay Watershed. The first two pages of the survey provide details about the Chesapeake Bay, the surrounding Watershed, the effects of nutrient and sediment pollution, and questions about a respondent's use of the Chesapeake Bay and lakes in the Watershed. Respondents are then introduced to the attributes that appear in the choice questions and are given the current attribute levels, and future baseline conditions in 2025. This paper focuses on the "constant baseline" version of the survey, in which the status quo option specifies environmental conditions that remain unchanged in the absence of additional pollution reductions.[4]

The survey then describes the practices that would improve conditions and how the costs of these practices would be passed onto households through an increased cost of living. Immediately preceding the choice questions are instructions describing the referendum and "cheap talk" script (Cummings and Taylor 1999) that instructs respondents to vote as if their household would actually face the costs shown and reminds them of their household budget constraint.

The DCE questions present status quo and policy scenarios represented by attributes describing environmental conditions in the year 2025 and household costs for each option. Through focus groups and consultation with experts on the ecology of the Chesapeake Bay and Watershed, we identified the most salient environmental attributes that are expected to change as

[4] Versions of the survey specifying status quo conditions that are worse than current conditions or better than current conditions were sent to households in some geographic strata. Examination and comparison of these data are the subject of future research.

a result of the reduced nutrient and sediment loadings entering the Bay. In defining these attributes it was also important that they could be linked to quantitative models that predict future changes under the baseline and TMDL scenarios. We limited the set of potential attributes to those that are likely to enter household utility functions directly, rather than inputs into an ecological production function (Boyd and Krupnick 2009, 2013). The attributes, including household costs, and possible attribute levels are shown in Table 1[5].

Table 1. Attributes and Attribute Levels

Attribute	**Attribute Levels**	
	Status quo	Policy Options
Water Clarity (feet)	3	3; 3.5; 4.5
Adult Striped Bass (millions)	24	24; 30; 36
Adult Blue Crab (millions)	250	250; 285; 328
Oysters (tons)	3,300	3,300; 5,500; 10,000
Low Algae Level Lakes	2,900	2,900; 3,300; 3,850
Annual Household Cost	$0	$20; $40; $60; $180; $250; $500

The survey includes three choice questions. Figure 1 provides an example. Each question presents a status quo option with baseline attribute levels and zero cost and two policy options with some or all of the attributes improving and positive costs. Respondents are asked to vote for

[5] We recognize that rivers and streams are also affected by excess nutrients and sediment, however, lacking the ability to model the ecological impacts of reducing loads to rivers and streams and being limited in the number of attributes we can include in the choice questions we chose to confine freshwater benefits to improvements in lakes.

one of the three options in each choice question. Alternatives were constructed based on an orthogonal fractional factorial design, and then assigned alternative pairs that would reflect trade-offs at the margin (i.e., improvements in the attributes that are attained at the cost of decreases in other environmental attributes and/or increase of the overall cost of the "policy"). Finally, these pairs were grouped in such a way that variability of the environmental and cost attributes would be maximized within and across individual surveys.

Figure 1. Example Choice Question

	Conditions in 2025 (% change compared to today)		
Environmental Outcomes	**Option A**	**Option B**	**Option C**
Bay Water Clarity Average visibility	**3** feet (no change)	**3** feet (no change)	**3.5** feet (17% increase)
Striped Bass Adult Population	**24** million fish (no change)	**30** million fish (25% increase)	**24** million fish (no change)
Blue Crab Adult Population	**250** million crabs (no change)	**250** million crabs (no change)	**250** million crabs (no change)
Oysters Population	**3,300** tons (no change)	**5,500** tons (67% increase)	**3,300** tons (no change)
Watershed Lakes Lakes with <u>low</u> algae levels	**2,900** lakes (no change)	**3,300** lakes (14% increase)	**3,300** lakes (14% increase)
Your Cost of Living Permanent cost increase for your household starting next year	**$0 every year**	**$500 every year** or $41.67 every month	**$250 every year** or $20.83 every month
Your Vote Please mark <u>one</u> of the boxes to the right	Option A ☐	Option B ☐	Option C ☐

Following the choice questions is a set of debriefing questions to identify factors that affect respondents' choices. In some cases responses are used as control variables in the estimation equation; in other cases responses are used to identify invalid responses to the choice questions.

Finally, the survey collects demographic data from the respondents in order to compare the sample with the population and to provide conditioning variables for WTP estimation.

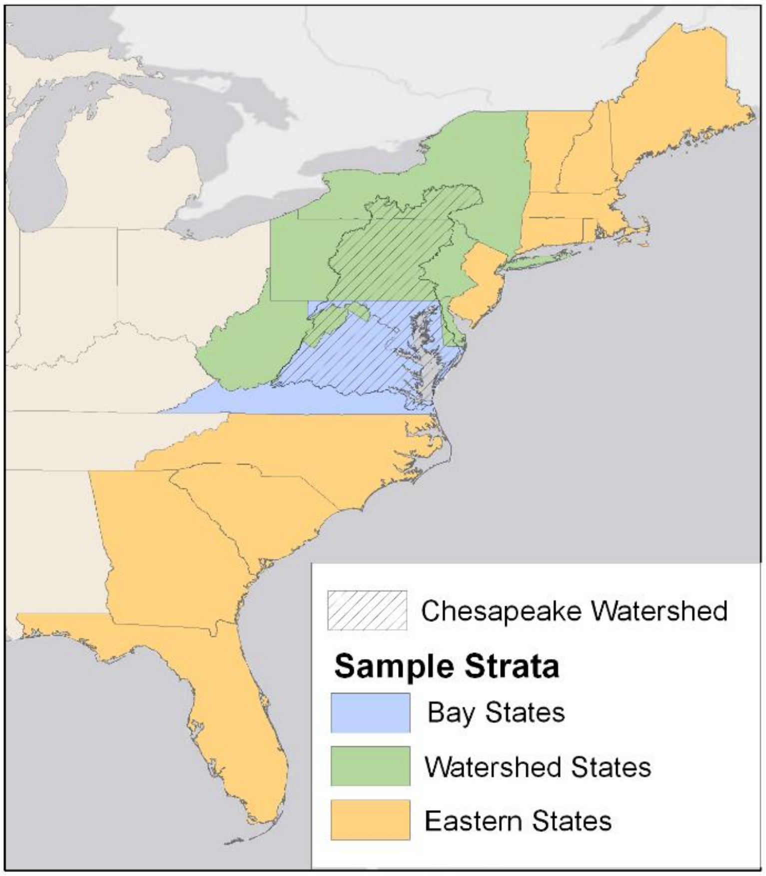
4. SAMPLE FRAME AND DATA

Following an extensive pre-test, the survey was administered via mail to a random sample of individuals 18 years of age or older who reside in the District of Columbia or one of 17 US states that contain at least part of the Chesapeake Bay Watershed or lie within 100 miles of the US East Coast. The sample was stratified by geographic region based on proximity to the Chesapeake Bay and the surrounding Watershed, as indicated in Table 2 and Figure 2.

Table 2. Sample Stratification by State.

Strata	**States/Location**
Bay States	Maryland, Virginia, District of Columbia
Watershed States	Delaware, New York, Pennsylvania, West Virginia
Other East Coast States	Connecticut, Florida, Georgia, Maine, Massachusetts, New Hampshire, New Jersey, North Carolina, Rhode Island, South Carolina, Vermont

Figure 2. Map of Sample Frame and Geographic Strata.



The sample was allocated in equal proportions of 33% for each stratum, thus leading to the highest sampling rate in the Bay States stratum (which has the lowest population, about 5.4 million households), and the lowest sampling rate in the Other East Coast States stratum (which has the highest population, about 25.4 million households). The total sample size across all versions of the survey and three geographic strata was 6,600 households. We focus here only on the "constant" baseline version of the survey (see section 3), which was sent to 2,829 households, 943 in each of the three geographic strata.

We used the Tailored Design Method to implement the survey as a mail questionnaire (Dillman 2008). Five mailings were sent to each randomly selected household: a preview letter notifying them that they had been selected to participate in the survey; the survey was mailed one week later; a reminder postcard; a second survey was mailed to those who did not respond to earlier mailings; and a final reminder letter was mailed after the second survey. All letters indicated whether or not the household was located in the Watershed and described the purpose of the survey. As shown in Table 3 the Bay States have the highest response rate of 34.1%, followed by the Watershed States with 30.1% and Other East Coast States having the lowest response rate, 27.1%.[6] Among the 1,642 completed surveys, 671 were of the "constant" baseline version of the survey analyzed in this study.[7]

[6] To calculate the response rate we use the American Association of Public Opinion Research's Response Rate 3 (RR3) which accounts for the number of completed surveys and the number of eligible addresses. These response rates are calculated based on all three baseline versions of the survey. The focus of the current study is only on the constant baseline version of the survey, but since we do not know which survey versions were sent to ineligible addresses, we cannot distinguish response rates by survey versions. Given that the survey baseline versions were randomly assigned to households within each strata, and that the number of returned surveys across versions were very similar within a strata, we believe the response rates for just the constant baseline version of the survey are similar.

[7] "Constant" baseline versions of the survey were returned by 683 respondents, but twelve surveys were disregarded because participants did not respond to any of the three choice questions.

Table 3. Summary of Response Rates by Geographic Sampling Strata[a]

	Bay States	Watershed States	Other Eastern States	Overall
Surveys Mailed	2,828	1,868	1,868	6,600
Completed (C)	794	479	369	1,642
Refusals/unable (R)[b]	15	14	23	52
Undeliverable (X)[c]	174	92	150	414
Unknown eligibility (U)[d]	1,845	1,301	1,344	4,429
Eligibility rate (e)[e]	82.3%	84.3%	72.3%	80.4%
Response Rate[f]	34.1%	30.1%	27.1%	31.0%

a. Response rates calculated based on all three baseline versions of the survey.
b. Surveys that were returned blank or with a note declining to participate.
c. Undeliverable surveys (e.g., vacant, no such street address, no mail receptacle).
d. No information about final disposition of survey.
e. e=(C+R)/(X+C+R); reflects the portion of the surveys with known disposition that are delivered to a household.
f. RR=C/(C+R+eU); assumes the portion of the surveys with unknown eligibility is equivalent to the portion with known eligibility (i.e., some of the surveys with unknown eligibility are unlikely to have been delivered to a physical address; returning undeliverable mail items is one of the lowest priorities for the postal service).

The survey included several questions to probe the respondents' familiarity and experience with the Chesapeake Bay and freshwater lakes in the Watershed. Most of the sample had heard of the Chesapeake Bay (94%) and knew that excess nutrients and sediment could degrade water quality (80%). About a third of the sample reported visiting the Bay (35%) or lakes in the Watershed (32%) in the last 5 years for recreational purposes.

Table 4 summarizes the responses for debriefing questions that probe the respondents' attitudes related to environmental regulation as well as their comprehension of the choice task and information provided beforehand. When asked about the statement that "It is important to improve waters in the Chesapeake Bay no matter how high the cost," 38% of the respondents agreed (answering 4 or 5 on a Likert scale from 1=Strongly Disagree to 5=Strongly Agree). About 27% of respondents indicated they were against more regulations and government spending. Finally, 28% of respondents felt they should not have to pay to improve water quality in the Chesapeake Bay and Watershed lakes, and 38.6% of respondents disagreed with that statement.

Table 4. Responses to Attitudinal Debriefing Questions[a]

	Strongly Disagree 1	2	3	4	Strongly Agree 5	Don't Know
It is important to improve waters in the Chesapeake Bay no matter how high the cost	12.2%	10.9%	27.6%	19.1%	16.7%	4.6%
I am against any more regulations and government spending	21.9%	13.6%	23.0%	9.7%	16.8%	5.8%
My household should not have to pay to improve Bay waters and Watershed Lakes	20.7%	14.8%	24.7%	7.6%	19.2%	4.2%
I voted as if my household would actually face the costs shown	3.5%	1.9%	9.1%	15.1%	55.8%	4.7%
I voted as if the programs would actually achieve the results shown	4.0%	3.1%	11.0%	15.7%	50.0%	6.3%

a. Unweighted sample statistics based on full sample of n=671 respondents (unscreened) who received and returned a constant baseline version of the survey. Percentages do not necessarily sum to 100% due to some respondents skipping individual questions.

In order to avoid biases associated with protest responses, scenario rejection, and hypothetical bias, both the choice questions and debriefing questions were used to identify and remove respondents who exhibited such behaviors. Such respondents could bias the WTP estimates in either a positive or negative direction and therefore are excluded from the analysis. Respondents potentially exhibiting biases were identified as follows:

- *Protest* - Respondent always chose status quo in the choice questions, and agreed or strongly agreed with the statements “I am against any more regulations and government spending” and “My household should not have to pay any amount to improve Bay Waters and Watershed lakes.”
- *Warm-glow* – Respondent always chose most expensive option in the choice questions, and agreed or strongly agreed with the statement “It is important to improve waters in the Chesapeake Bay Watershed, no matter how high the costs.”
- *Hypothetical bias* – Respondent always chose a policy option in the choice questions, and disagreed or strongly disagreed with the statement “I voted as if my household would actually face the costs shown in the questions.”
- *Scenario Rejection* – Respondents who disagreed or strongly disagreed with the statement “I voted as if the programs would actually achieve the results shown by 2025.”

Table 5 displays the number of respondents identified under each of the screening criteria. The original unscreened sample includes 671 respondents who answered at least one of the choice questions. The main regression results and WTP estimates in this report are based on the 559 respondents who do not provide evidence of protest, warm- glow, scenario rejection, or hypothetical bias (see rightmost column in Table 5).

Table 5. Number of respondents under alternative screening criteria (Constant Baseline Survey)[a]

Number of Respondents	Unscreened	Warm-glow	Hypothetical Bias	Protest	Scenario Rejection	All Criteria
Eliminated	-	6	9	64	47	112
Remaining	671	665	662	607	624	559

a. Categories are not mutually exclusive. A respondent can potentially exhibit more than one type of bias.

Table 6 presents the characteristics of the final sample of respondents (n=559) used in the current study compared to the population. The sample is representative in terms of the proportion of respondents that are male, but does contain fewer minorities than the corresponding sample frame population. The sample also contains a higher proportion of people with at least a four-year college degree compared to the general population, though this difference is significant in only one of the three strata. Although not displayed here, in all strata the sample appears to somewhat under-represent individuals in the youngest age group, and over-represent individuals in the eldest group. The distribution of annual household income for the sample in each strata is compared to the corresponding population of households in Figure 3. The distribution of income is fairly similar, although the lowest income category appears under-represented in all three strata. In the final analysis any differences between the samples are later controlled for econometrically by conditioning WTP on these characteristics, and calculating welfare changes for a representative household.

Table 6. Sample and Population Demographic Comparisons.

	Main Survey	Sampling Stratum Population	t-test of means
Bay States Strata			
Male	50.5%	48.01%	0.70
Hispanic	5.15%	8.05%	-1.82*
Black	13.02%	26.10%	-5.37***
4-Year College Degree[a]	56.85%	53.21%	1.03
Watershed States Strata			
Male	49.48%	47.83%	0.46
Hispanic	6.63%	12.15%	-2.98***
Black	5.68%	14.71%	-5.15***
4-Year College Degree[a]	47.34%	43.66%	1.01
Other Eastern States Strata			
Male	47.13%	48.04%	-0.23
Hispanic	4.83%	13.56%	-4.89***
Black	8.22%	18.34%	-4.44***
4-Year College Degree[a]	54.78%	43.63%	2.78***

*** p<0.01, ** p<0.05, * p<0.1

a. Only includes respondents over age 25 in order to match to corresponding population statistic.

Figure 3. Annual Household Income by Strata

Sample percentages based on the 509 respondents (out of 559) that answered the income question (192 in the Bay States stratum, 172 in the Watershed States stratum, and 145 in the Other Eastern States stratum).

5. ANALYSIS AND RESULTS

Empirical Model

The empirical model is based on random utility theory, which posits that utility is composed of a deterministic component, $v(\cdot)$, and an unobserved random component ε. The utility u_{ij} experienced by household *i* from alternative *j* is defined by the conditional indirect utility function:

$$u_{ij} = v(\boldsymbol{x_j}, Y_i - C_j) + \varepsilon_{ij} = v_{ij} + \varepsilon_{ij} \qquad (1)$$

where $v(.)$ or v_{ij} is the deterministic component of utility, and is a function of a vector of attributes describing the alternative $\boldsymbol{x_j}$, as well as numeraire consumption, Y_i - C_j (income minus the cost of alternative *j*). The vector $\boldsymbol{x_j}$ includes attributes of the Bay (i.e., water clarity, bass, crab, and oyster populations) and the number of Watershed lakes with low algae levels. Utility also depends on a stochastic component ε_{ij} that is not observable.

The literature offers no clear guidance regarding the choice of specific functional forms for $v(\cdot)$ within choice experiment estimation. In practice, linear forms are often used (Johnston et al., 2003), although some have applied more flexible forms to allow for nonlinearities over the attribute space (Cummings et al., 1994). We adopt a log-linear model that captures diminishing marginal utility while preserving more degrees of freedom than a model with higher order effects.[8] Omitting the choice question subscript *t* for notational ease the most basic model is:

[8] Preliminary analysis of the data compared linear, quadratic and log-linear specifications of the indirect utility function. Coefficient and WTP estimates were stable across specifications and the quadratic specification showed that marginal utility is diminishing in some attributes.

$$u_{ij} = \boldsymbol{\beta} ln(\boldsymbol{x}_j) + \gamma C_j + \varepsilon_{ij} \quad (2)$$

where γ is the negative of the marginal utility of income.

When estimating the empirical models, we include an indicator variable (0 or 1) for the status quo option to estimate a constant SQC_{ij}. that captures respondents' tendencies to prefer the status quo regardless of the attribute improvements. A positive and statistically significant *SQC* would suggest respondents tend to favor the status quo (perhaps reflecting protest responses) while a negative and significant *SQC* would suggest that respondents favor a policy option in general, but not necessarily due to specified attribute levels. Such behavior could be due to respondents considering omitted factors (i.e., improvements to aspects of the environment that were not described in the choice alternatives), or a general warm-glow for doing something to help the environment. Omission of *SQC* in the estimating equation could confound estimates of $\boldsymbol{\beta}$ and γ. To account for heterogeneity across respondents with regards to both possibilities, the models are estimated following a mixed logit specification, where *SQC* is allowed to vary across respondents following an assumed normal distribution, $SQC_{ij} \sim N(\mu_{SQC}, \sigma^2_{SQC})$ so that the utility respondent *i* derives from alternative *j* is

$$u_{ij} = \boldsymbol{\beta} ln(\boldsymbol{x}_j) + \gamma C_j + SQC_{ij} + \varepsilon_{ij} \quad (3)$$

where the parameters to be estimated are $\boldsymbol{\beta}$, γ, μ_{SQC}, σ^2_{SQC}. Assuming ε follows a type I extreme value (Gumbel) distribution, the model can be estimated as conditional logit and mixed logit (Maddala 1983; Greene 2003; Train 2009) and the probability that person *i* chooses alternative *j* becomes

$$P\left(j \mid x_{k,} C_k\right)=\frac{\exp\left(\boldsymbol{\beta} ln\left(\boldsymbol{x}_j\right)+\gamma C_j+SQC_{ij}+\varepsilon_{ij}\right)}{\sum_k \exp\left(\boldsymbol{\beta} ln\left(\boldsymbol{x}_k\right)+\gamma C_k+SQC_{ik}+\varepsilon_{ik}\right)} \tag{4}$$

The coefficients in equation (4) are estimated as a mixed logit model where *SQC* is the only stochastic coefficient and the other coefficients are treated as fixed.[9]

Marginal willingness to pay (MWTP) for attribute x_k given a reference level $\widetilde{x_k}$ is:

$$MWTP(x_k)=\frac{\partial u_{ij} / \partial \boldsymbol{x_{kij}}}{-\partial u_{ij} / \partial C_{ij}}=\frac{\boldsymbol{\beta}}{-\gamma \widetilde{\boldsymbol{x_k}}}. \tag{5}$$

Equations (1) – (5) implicitly assume that the marginal utilities corresponding to the environmental attributes and income are constant across households, something we subsequently relax by adding a series of interaction terms between household characteristics and the environmental and cost attributes[10]. Such characteristics include whether respondents are recreational users or non-users of the Chesapeake Bay and lakes in the Watershed, annual household income, and other demographic characteristics.

[9] Mansfield et al. (2012) took a similar approach in their SP study of dam removal in the Klamath River Basin.

[10] The inclusion of such interaction terms is a common approach in the literature to account for observed respondent heterogeneity (e.g., Mansfield et al., 2012; Van Houtven et al., 2014).

Sample Weights

Sampling intensities and response rates differ across the strata; both are lower in strata located farther from the Chesapeake Bay. In order to account for these differences we pool data across all strata, weighting responses to ensure households in each of the strata are equally represented.

Observations are weighted according to the inverse of the probability of a household being included in the final estimation sample. This probability depends on the household being sent a constant baseline version of the survey, returning that survey and completing the choice questions, and whether or not they were screened from the sample. The sample weights applied to the 559 respondents are shown in Table 7.

Table 7. Sample Weights across Geographic Strata

Stratum	Total Households	Households in Estimation Sample	Base Weight
Bay States	5,479,176	206	26,598
Watershed States	13,442,787	195	68,937
Other Eastern States	25,431,478	158	160,959

Regression Results

Equation (3) is estimated using simulated maximum likelihood with 200 Halton draws (Train, 2009).[11] Since each respondent faces three choice tasks, the panel structure of the data is accounted for and the standard errors clustered at the respondent level.

The base RUM regression coefficient results (Model 1) are presented in column 1 of Table 8. The coefficients corresponding to water clarity, populations of bass, crab, oysters, and the number of Watershed lakes with low algae are all positive and statistically significant, meaning that the marginal utility for an improvement in these attributes is positive. The negative and significant coefficient on the cost attribute implies a positive marginal utility of income, as expected.

Table 8. RUM Regression Coefficient Results.[a]

Variables	Model 1 (Base Model)	Model 2	Model 3	Model 4
ln(clarity)	0.9356**	0.3933	0.4746	0.5994
	(0.4586)	(0.5545)	(0.5821)	(0.7884)
ln(bass)	1.1785**	1.2657**	1.3516**	3.1766***
	(0.5088)	(0.5946)	(0.6380)	(0.9129)
ln(crab)	2.2465***	2.3412***	2.5064***	1.5940
	(0.6755)	(0.8141)	(0.8828)	(1.3704)
ln(oysters)	0.4189**	0.2875	0.4139**	0.7259***
	(0.1655)	(0.1884)	(0.1922)	(0.2697)
ln(lakes)	3.9673***	3.8423***	3.9164***	2.7010**
	(0.6350)	(0.7278)	(0.7419)	(1.0818)

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

[11] The random utility models are estimated in Stata 14 using the "mixlogit" command (Hole 2007). To ensure convergence was reached the same models were estimated using mixed logit software for Matlab developed by Kenneth E. Train and available for download here: http://eml.berkeley.edu/~train/software.html.

Cost	-0.0074***	-0.0075***		
	(0.0007)	(0.0007)		
cost × income 0-50k			-0.0065***	-0.0093***
			(0.0012)	(0.0016)
cost × income 50-100k			-0.0094***	-0.0130***
			(0.0012)	(0.0017)
cost × income 100k+			-0.0068***	-0.0102***
			(0.0012)	(0.0018)
status quo (mean)	-1.5335***	-1.5463***	-1.6706***	-1.8676***
	(0.3551)	(0.3545)	(0.3855)	(0.4377)
status quo (std dev)	4.2770***	4.2156***	4.3323***	4.3451***
	(0.4705)	(0.4623)	(0.5077)	(0.5000)
ln(clarity) × bay user		1.9440**	1.8581**	0.8432
		(0.8816)	(0.9332)	(1.0135)
ln(bass) × bay user		-0.1524	-0.1610	-0.6487
		(0.9350)	(1.0089)	(1.1440)
ln(crab) × bay user		-0.5496	-0.5812	-1.9206
		(1.1712)	(1.2436)	(1.4558)
ln(oysters) × bay user		0.5469	0.4709	0.4969
		(0.3330)	(0.3450)	(0.4159)
ln(lakes) × lake user		0.5742	0.4706	0.4973
		(1.2481)	(1.3246)	(1.4541)
ln(clarity) × degree				1.4712
				(0.9793)
ln(bass) × degree				-0.8196
				(1.0857)
ln(crab) × degree				3.9350**
				(1.6127)
ln(oysters) × degree				-0.1228
				(0.3601)
ln(lakes) × degree				3.2156**
				(1.2933)
cost × degree				0.0021
				(0.0016)
ln(clarity) × black				-2.9671
				(2.3431)
ln(bass) × black				-5.6707***
				(1.3625)

ln(crab) × black				-3.4157
				(2.6455)
ln(oysters) × black				-1.4508**
				(0.6128)
ln(lakes) × black				-3.6292
				(2.6340)
cost × black				0.0092***
				(0.0021)
ln(clarity) × hispanic				-5.7168*
				(3.1388)
ln(bass) × hispanic				-2.9235
				(2.3688)
ln(crab) × hispanic				-8.5809***
				(2.6105)
ln(oysters) × hispanic				-1.4672*
				(0.8659)
ln(lakes) × hispanic				0.9446
				(3.8028)
cost × hispanic				0.0072***
				(0.0017)
Observations	4,719	4,719	4,308	3,888
ll	-1.0702e+08	-1.0660e+08	-9.6853e+07	-8.3023e+07

*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

a. Standard errors in parentheses, clustered at the respondent-level. Mixed logit regressions estimated using 200 Halton draws.

The *SQC* is allowed to vary randomly across respondents, but the negative and statistically significant mean value (*status quo*) shows that, on average, respondents are more likely to choose a policy option (Option B or Option C) over the status quo (Option A), even after controlling for the attribute improvements. The statistically significant estimated standard deviation corresponding to the *SQC* shows that there is variation across respondents regarding the *SQC*, supporting the mixed logit specification. The *SQC* is fairly large, perhaps because of general concerns about the future of the environmental quality in the Chesapeake Bay Watershed. The magnitude of the *SQC* makes it all the more important to control for these effects, as we do, when later estimating household WTP for TMDL-related improvements.

In Model 2, interaction terms between the Bay attributes (clarity, bass, crab, and oysters) and a dummy variable denoting active users of the Bay are added, as well as an interaction term between the Watershed lakes attribute and a dummy variable identifying users of Watershed lakes. Users are identified as anyone who reports participating in recreational activities at the Bay or Watershed lakes within the last five years. Only the coefficient estimate for the interaction term between bay users and water clarity is significant, indicating that users hold a premium for bay water clarity. The positive coefficients on the un-interacted environmental attributes, when significant, suggest that non-users hold a positive marginal utility, namely for bass, crab, and Watershed lakes.

Exploratory models not reported here contained interaction terms between the environmental attributes and a dummy variable denoting whether the household is located within the Watershed (hatched area shown in Figure 2). One might expect that residents within the Watershed have stronger preferences for improvements in the Chesapeake Bay and Watershed lakes because they are more familiar with these resources and are more likely to be users. Consistently we found that, conditional on user status, the coefficients on these in-Watershed interaction terms were not significant, neither individually nor jointly. Similar results were found using finer measures of proximity to the Bay, including: linear distance (kilometers), inverse distance, the natural log of distance, a dummy variable denoting respondents within 50 kilometers, and variables denoting the geographic strata. The lack of a WTP distance gradient has been found in other stated preference studies (e.g., Johnston and Ramachandran, 2014; Johnston et al., 2015; Rolfe and Windle, 2012), and is not surprising given the iconic nature of the Chesapeake Bay and the potentially large nonuse values people may hold.[12]

Following equation (5), the annual MWTP estimates from Models 1 and 2 are calculated and presented in Table 9. Since utility is non-linear in the attributes, MWTP is a function of the attribute levels. The values shown in Table 9 are evaluated at the status quo levels for each attribute. Notice that as shown in equation (6), the *SQC* is not included when calculating MWTP, and so the following estimates are not biased by the average tendency to favor a policy option irrespective of changes in the environmental attributes. All WTP estimates are reported in 2014 dollars.[13]

Looking first at Model 1, we see that the MWTP estimates are positive and significant for all attributes. Estimated annual WTP for a one foot increase in Bay water clarity is $42.06. The relatively large estimate for clarity is due mainly to the units used to express changes in that attribute on the survey. A one foot increase in water clarity is a large improvement compared to a one unit increase in the other environmental attributes. Respondents are willing to pay $6.62 per year to increase the population of striped bass by one million, $1.21 for one million additional blue

[12] Mansfield et al (2012) found an inverse distance gradient for a stated preference study of the Klamath River Basin.
[13] Standard errors for all WTP estimates are obtained via the "delta" method. Earlier results were similar when using bootstrapped standard errors over 1,000 iterations.

crabs, and about $0.02 for a one ton increase in oysters. Finally, an additional low algae lake is valued at $0.18 per respondent per year.

Table 9. Marginal Willingness to Pay (MWTP) Estimates (2014 dollars)

	Model 1	Model 2	
		Users	Nonusers
Clarity (feet)	42.0641**	103.3865***	17.3973
	(20.3675)	(32.0027)	(24.3824)
Bass (million fish)	6.6230**	6.1556	6.9984**
	(2.7697)	(4.3016)	(3.2321)
Crab (million crab)	1.2121***	0.9510**	1.2427***
	(0.3638)	(0.4819)	(0.4351)
Oysters (tons)	0.0171***	0.0336***	0.0116
	(0.0065)	(0.0111)	(0.0075)
Lakes (# of lakes)	0.1845***	0.2021***	0.1758***
	(0.0330)	(0.0485)	(0.0373)

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

Because we included Bay and lake user interaction terms, Model 2 allows for the comparison of MWTP between users and non-users. Users of the Bay hold a significantly higher MWTP of $103.39 per year for a one foot improvement in clarity compared to non-users for whom MWTP is not significant. This result is consistent with earlier focus group and cognitive interview results in which water clarity was consistently the most important attribute for users of the Bay. The MWTP values are fairly similar for crab and lakes between users and non-users. The point estimates are also similar for bass and oysters, although the former is statistically significant only among non-users and the latter among users.[14]

Building off of the previous specifications, Models 3 and 4 in Table 8 further account for household heterogeneity by interacting the cost and/or environmental attributes with household income, education, race, and ethnicity. In Model 3 the marginal utility of income is allowed to vary across income groups. Respondents were divided into three annual income categories ($0-50k, $50-100k, and >$100k), and dummy variables denoting each were interacted with *cost.* In theory, one may expect these coefficients to be decreasing in absolute terms, reflecting a decreasing marginal utility of income. However, the middle income group has the highest marginal utility of income although it is not statistically different across the groups.[15]

Additional interaction terms are added in Model 4 to account for household heterogeneity based on education, race, and ethnicity. Regarding education, we add interactions between cost and the environmental attributes with a dummy variable denoting respondents who have a Bachelor's degree or higher (*degree*). The coefficients corresponding to the interaction terms between *degree* and the environmental attributes are mixed, but when significant are positive, suggesting that relatively educated respondents hold a higher marginal utility for some attributes,

---

[14] A nonlinear Wald test rejects the null hypothesis that the MWTP for clarity is statistically equal across users and nonusers ($\chi^2(1)=4.63$, p=0.0313). Similar tests reveal no statistically significant difference in MWTP for bass, crab, and Watershed lakes. Although one can reject the null hypothesis that users and nonusers have an equal MWTP for oysters at the $p<0.10$ level ($\chi^2(1)=2.81$, p=0.0937).

[15] A Wald test fails to reject the null hypothesis that the coefficients on the cost interactions are equal across the three income categories ($\chi^2(2)=3.83$, p=0.1472). In earlier analysis several alternative models were explored, such as including interactions between *cost* and a linear 1 to 7 scalar denoting the different income categories, and a dummy variable denoting households with income levels above the median. These models yielded no significant evidence suggesting that the cost coefficient varies with income.

all else constant. Wald tests also reject the null hypothesis that these coefficients are all equal to zero ($\chi^2(6)=16.29$ $p = 0.0123$), and so the coefficients on the *degree* interaction terms are jointly significant.

The interaction terms between the environmental attributes and dummy variables denoting whether a respondent is Black or African American (*black*) are all negative, and are statistically significant for the bass and oyster interaction terms. Similarly, interactions between the environmental attributes and a dummy variable denoting respondents of Hispanic, Latino, or Spanish origin (*hispanic*) are also often negative, and significant for bay water clarity, crab, and oysters. The interactions between *cost* and *black* or *hispanic* are positive and significant, suggesting that such respondents hold a lower marginal utility of income, all else constant. The coefficients on these interaction terms are jointly, and often individually significant, suggesting that race and ethnicity are important factors in estimating a household's WTP.[16, 17]

Next, models 1 through 4 are re-estimated with additional interaction terms to account for some respondents that may have considered improvements in attributes that were not specified in the choice experiment. A general concern with stated choice experiments is that respondents may be thinking of other factors when answering the choice questions. If respondents correlate their

---

[16] Wald tests of the null hypotheses that the coefficients on these interaction terms are all jointly equal to zero were rejected for both the *black* and *hispanic* interaction terms ($\chi^2(6)=48.85$, p=0.0000 and $\chi^2(6)=61.69$, p=0.0000, respectively).

[17] Additional models not reported here explored other aspects of household heterogeneity. These models included interaction terms between the environmental and cost attributes and the respondent's age, whether they are male, and whether the household has any children under the age of 18. The resulting coefficients were statistically insignificant.

perceptions of changes in such factors with one of the described attributes, then an omitted variable bias could arise. Although such considerations could be valid in reality, how these factors change under each alternative was not expressed to respondents, and in the absence of a clearly and quantitatively described mapping, respondents may assert their own beliefs, which could be inaccurate and are unknown to the researcher (Boyd and Krupnick, 2013; Johnston et al., 2013). In the following four models the potential for such biases is examined.

In Table 10, models 1' through 4' are the same as those presented above, but now include additional interaction terms denoting respondents who stated that they considered omitted factors when responding to the choice questions. An interaction term between the lakes attribute and a dummy variable (*not Watershed lakes*) denoting respondents who agreed or strongly agreed that "water quality improvements to lakes outside the Chesapeake Bay Watershed" affected their vote was added to the right-hand side of the regression. Respondents were explicitly told lakes outside the Watershed would not be affected by new programs, however, some respondents may not have fully absorbed this information in light of the number who agreed with the statement in this debriefing question (50.3%). Similarly, a series of interaction terms between *bass*, *crab*, and *oysters* with a dummy variable *food*, which denotes respondents who agreed or strongly agreed that "changes in the quality or price of seafood" affected their vote, was also added. The results of the re-estimated RUM regressions are presented in Table 10.

Table 10. RUM Regressions with Interaction Terms denoting Respondents Who Considered Omitted Factors[a]

Variables	Model 1'	Model 2'	Model 3'	Model 4'
ln(clarity)	0.9321**	0.3812	0.4590	0.5207
	(0.4551)	(0.5476)	(0.5743)	(0.7739)
ln(bass)	0.5176	0.6080	0.7469	2.3648**
	(0.6265)	(0.6728)	(0.7305)	(0.9215)
ln(crab)	1.4444*	1.5459	1.3458	0.5145
	(0.8462)	(0.9519)	(1.0031)	(1.4887)
ln(oysters)	0.2675	0.1456	0.3089	0.6373**
	(0.2095)	(0.2297)	(0.2306)	(0.2841)
ln(lakes)	2.1433**	2.0799**	2.2857**	1.0201
	(0.8553)	(0.9063)	(0.9803)	(1.2417)
ln(bass) × food	1.3901	1.3653	1.1059	1.7602
	(0.9206)	(0.9159)	(0.9821)	(1.0826)
ln(crab) × food	1.7153	1.7358	2.6384*	3.0557**
	(1.2678)	(1.2710)	(1.3643)	(1.4405)
ln(oysters) × food	0.3801	0.3771	0.2568	0.2299
	(0.3023)	(0.3008)	(0.3161)	(0.3446)
ln(lakes) × not Watershed lakes	3.2142***	3.2195***	2.9625**	3.0104**
	(1.1769)	(1.1797)	(1.2091)	(1.2916)
Cost	-0.0075***	-0.0076***		
	(0.0007)	(0.0007)		
cost × income 0-50k			-0.0065***	-0.0093***
			(0.0012)	(0.0016)
cost × income 50-100k			-0.0093***	-0.0130***
			(0.0012)	(0.0017)
cost × income 100k+			-0.0069***	-0.0104***
			(0.0011)	(0.0018)
status quo (mean)	-1.5013***	-1.5156***	-1.6315***	-1.8393***
	(0.3422)	(0.3417)	(0.3721)	(0.4243)
status quo (std dev)	3.9883***	3.9314***	4.0550***	4.1104***
	(0.4416)	(0.4326)	(0.4772)	(0.4703)
ln(clarity) × user		1.9931**	1.8829**	0.8154
		(0.8773)	(0.9284)	(1.0155)
ln(bass) × user		-0.1366	-0.1400	-0.5942
		(0.9466)	(1.0119)	(1.1624)

ln(crab) × user	-0.5956	-0.6221	-2.0087
	(1.1623)	(1.2383)	(1.4688)
ln(oysters) × user	0.5152	0.4502	0.4517
	(0.3294)	(0.3415)	(0.4152)
ln(lakes) × user	0.2635	0.2492	0.3678
	(1.2560)	(1.3099)	(1.4308)
ln(clarity) × degree			1.6467*
			(0.9760)
ln(bass) × degree			-0.9623
			(1.1022)
ln(crab) × degree			3.6810**
			(1.6141)
ln(oysters) × degree			-0.1192
			(0.3660)
ln(lakes) × degree			3.3133***
			(1.2774)
cost × degree			0.0021
			(0.0016)
ln(clarity) × black			-3.0730
			(2.2527)
ln(bass) × black			-5.7144***
			(1.3910)
ln(crab) × black			-3.7001
			(2.7024)
ln(oysters) × black			-1.3033**
			(0.5960)
ln(lakes) × black			-2.9820
			(2.6213)
cost × black			0.0089***
			(0.0021)
ln(clarity) × hispanic			-5.4383*
			(3.0483)
ln(bass) × hispanic			-3.1546
			(2.1119)
ln(crab) × hispanic			-8.6611***
			(2.4099)
ln(oysters) × hispanic			-1.4052*
			(0.8217)

ln(lakes) × hispanic				0.4350
				(3.5679)
cost × hispanic				0.0072***
				(0.0017)
Observations	4,719	4,719	4,308	3,888
ll	-1.0622e+08	-1.0580e+08	-9.6171e+07	-8.2290e+07

*** p<0.01, ** p<0.05, * p<0.1.
[b] Standard errors in parentheses, clustered at the respondent-level. Mixed logit regressions estimated using 200 Halton draws.

The interaction terms between *bass*, *crab*, and *oysters* with *food* are insignificant in most models. The only exceptions are that the coefficient on *ln(crab)* × *food* is significant in models 3' and 4' (at the p=0.10 and p=0.05 level, respectively). This provides some weak evidence that respondents who may have considered changes in seafood price or quality hold a premium for improvements in crab populations. Across all four models the coefficients on *ln(lakes)* × *not Watershed lakes* are positive and significant, suggesting that respondents who said they were considering lakes outside the Watershed when answering the choice questions are willing to pay more for lake improvements than the rest of the sample.

It is unclear whether such influences reflect omitted variable bias, differing cognitive abilities to parse WTP for water quality improvements to certain water bodies as distinguished from water bodies in general, or actual preference heterogeneity among households. In focus group and cognitive interviews involving earlier versions of the survey instrument, we found evidence that some respondents were answering the debriefing questions in general terms, and were not necessarily responding with their thought processes for the choice questions in mind. For this reason, hypothetical bias, warm-glow, and protest responses were identified based on responses to both the debriefing questions and the choice questions. Unfortunately, there is no clear way to incorporate the latter when identifying respondents who may actually have considered omitted factors when making their choices. At the same time, respondents who agreed with these statements may also have systematically stronger preferences towards improvements in Watershed lakes and populations of fish and shellfish in the Bay. In either case, such influences can be controlled for econometrically and, if desired, can be excluded from the resulting WTP estimates (by excluding the coefficients corresponding to the *food* and *not Watershed lakes* interaction terms when calculating WTP).

Since it remains unclear whether the coefficients on these interaction terms capture omitted variable biases, valid preference heterogeneity, or both, we use both sets of models (Models 1 through 4 and 1' through 4') to calculate WTP for the environmental improvements projected under the TMDL.

6. NON-RESPONSE BIAS STUDY

In order to determine if and how respondents and non-respondents differ, EPA conducted a non-response bias study (NRBS) in which a short survey was administered to a random sample of households that received the main survey but did not complete and return it. The short questionnaire asks awareness, attitudinal, and demographic questions that can be used to statistically examine differences between respondents and non-respondents.

Previous survey research shows prepaid financial tokens are one of the greatest contributions to an increased response rate (Dillman 2008). It has been demonstrated that a financial token may pull in responders that otherwise are not interested in participating in the survey (Groves et al. 2006), an issue that is of particular relevance to non-response bias. EPA included $2 in cash as an unconditional incentive for completion of the short questionnaire to encourage response from this population. The resulting response rate was 19% (AAPOR RR3) with 263 non-response follow up questionnaires completed and returned. The subset of the questions from the main questionnaire selected for the non-response bias study survey is discussed below.

Familiarity with the Chesapeake Bay Watershed and Watershed Issues. After a brief introduction, four questions are presented that inquire about individuals' awareness and use of the Chesapeake Bay and Watershed lakes. It is likely that awareness and use of an environmental commodity are correlated with individuals' willingness-to-pay (WTP) for improvements (e.g., Johnston et al. 2005). It is therefore important to assess whether there are systematic differences in these responses across respondents to the main survey and those to the non-response follow-up questionnaire. Table 11 compares the percentage of positive responses to these questions.

Table 11. Comparison of Familiarity with the Chesapeake Bay Watershed and Watershed Issues[a]

	Main Survey	NRBS	t-test of Means
Before receiving the survey, had you heard of the Chesapeake Bay?	94%	86%	3.17***
On average, how often do you see the Chesapeake Bay?			
Never	35%	40%	1.45
Less than once a month	41%	34%	1.84*
More than once a month	17%	14%	1.41
On average, how often do you see Watershed Lakes?			
Never	35%	41%	1.62
Less than once a month	31%	29%	0.54
More than once a month	21%	13%	3.35***
In the last five years, have you participated in recreational activities (including swimming, boating, fishing, or viewing nature) at the Chesapeake Bay?	38%	32%	1.59
In the last five years, have you participated in recreational activities (including swimming, boating, fishing, or viewing nature) at Watershed Lakes?	36%	30%	0.78
Before taking this survey, were you aware that too much nutrients or sediment can degrade water quality?	79%	73%	1.83*

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

[a] Main survey sample statistics are based on the unweighted screened sample of 559 constant baseline surveys; NRBS sample statistics are based on the full unweighted sample of 263 completed questionnaires.

Attitudes towards Environment and Regulations. We can assess whether non-respondents did not complete the main survey for reasons related to the survey topic by comparing responses to these questions about attitudes toward water quality improvements in the Chesapeake Bay

Watershed, costs to one's household, and government regulations across the main survey study and the non-response bias study. Table 12 shows the mean response to each question (on a scale of 1 = Strongly Disagree to 5 = Strongly Agree) and the results of Mann-Whitney tests comparing the distributions of responses across samples.[18] The only statement for which the null hypothesis that the samples are drawn from the same population can be rejected is "My household should not have to pay to improve Bay Waters and Watershed Lakes," with a larger proportion of the NRBS sample in agreement.

Table 12. Comparison of Attitudinal Responses[a]

	Main Survey	NRBS	Mann-Whitney U test
It is important to improve waters in the Chesapeake Bay Watershed, no matter how high the costs	3.20	3.63	-0.799
I am against anymore regulations and government spending	2.84	2.89	-1.419
My household should not have to pay any amount to improve Bay Waters and Watershed Lakes	2.89	3.47	2.859***
It is difficult for me to find time to take surveys	2.71	3.26	1.524

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

[a] Main survey sample statistics are based on the unweighted screened sample of 559 constant baseline surveys; NRBS sample statistics are based on the full unweighted sample of 263 completed questionnaires.

Demographics. By including demographic questions in both the survey and non-response follow-up survey, comparisons of household characteristics can be made across the samples of responding

[18] Also known as a Wilcoxon rank-sum test. See Hogg et al. (2005, page 541) for details.

and non-responding households. These data can also be compared to household characteristics from the sample frame population, which are available from the 2010 Census.

Table 13. Demographic Comparison[a]

	Main Survey	Non-response Follow up	Sample Frame Population[b]
Male	48%	50%	48%
Hispanic	5%	9%	13%
Black	8%	13%	18%
Pacific Islander	0%	4%	<1%
Asian	4%	8%	5%
White	88%	78%	73%
4-Year College Degree[c]	53%	44%	45%
Income			
Under $25,000	14%	17%	24%
$25,000-$49,999	19%	16%	24%
$50,000-$74,999	17%	13%	18%
$75,000-$99,999	16%	10%	12%
$100,000-$149,999	12%	11%	12%
$150,000-$199,999	6%	5%	5%
$200,000 or more	7%	8%	5%

[a] Main survey sample statistics are based on the screened sample of 559 constant baseline surveys; NRBS sample statistics are based on the full sample of 263 completed follow-up questionnaires. The sample statistics are weighted to account for the stratified random sampling. Based on the 2010 US Census, strata weights are equal to the number of households within each geographic strata (see Table 7), divided by the total number of households in the study frame (44,353,441). Source: US Census Bureau American Fact Finder

[b] Only includes individuals over 25 years of age to match population statistic derived from 2010 American Communities Survey.

Data from the main survey and the NRBS suggest there is some potential for non-response bias. People who are familiar with the Chesapeake Bay and were aware of nutrient and sediment pollution before receiving the main survey were more likely to complete and return it than the NRBS sample. While only a few of these differences are statistically significant, respondents' experiences could influence their responses to the policy choice questions and our WTP estimates. Data from the attitudinal questions, however, are less conclusive. The two samples have very similar attitudes toward more regulation and government spending. NRBS respondents are more likely to agree that water quality should be improved regardless of cost but are also more likely to say that their household should not have to pay for those improvements.

The demographic characteristics of both samples were compared to the population of the sample frame. With respect to many demographic characteristics (e.g., proportion of individuals that identify themselves as Hispanic, Black, White, or as having a 4-year college degree or higher) the NRBS sample was closer to the population than the main survey sample. For this reason when extrapolating WTP we use the NRBS proportions when conditioning WTP estimates on variables where data on the broader population are not available; such as a proportion for users of the Bay and Watershed lakes. When data are available, the extrapolated WTP estimates are conditioned on the characteristics of the sample frame population.

7. HOUSEHOLD AND TOTAL WTP

We use the preference parameter estimates from the RUM regressions to estimate the benefits of the environmental improvements projected to occur under the full implementation of the TMDL, incremental to the projected baseline scenario. The projected changes for the attributes included in the choice questions are presented in Table 14.

Table 14. Projected Changes in Environmental Attributes

Environmental Attribute	Baseline Level	Improvement	Percent Improvement
Bay Water Clarity	3 feet	+0.361 feet	+12 %
Striped Bass Populations	24 million fish	+1.032 million fish	+4.3 %
Blue Crab Populations	250 million crabs	+41 million crab	+16.4 %
Oyster Populations	3,300 tons	+541.2 tons	+16.4 %
Low Algae Watershed Lakes	2,900 lakes	+455 lakes	+15.7 %

Improvements in water clarity from the TMDL come from data provided by the US Environmental Protection Agency's (EPA) Chesapeake Bay Program Office (CBPO). The Chesapeake Bay Estuary Model projected spatially explicit water clarity levels across the Bay and tidal tributaries under two scenarios: (i) a baseline accounting for actions undertaken before the TMDL was enacted and future actions that would have been implemented in the absence of the TMDL, and (ii) a policy scenario in which the TMDL is fully implemented according to the Phase II watershed implementation plans (WIPs) put forth by each of the six States and the District of

Columbia.[19, 20] According to the Estuary Model, the largest gains in water clarity are expected in the upper Bay and tidal tributaries, with smaller improvements occurring closer to the mouth of the Bay. We use an average improvement across the Bay and tidal tributaries.

The changes in striped bass, blue crab, and oyster populations are based on a summary of expert judgments from a panel of six water quality and fishery experts assembled at an EPA workshop held in January 2013. The experts were asked, during the course of facilitated open discussion, to provide best professional judgments of changes in species stock sizes with the TMDL relative to current conditions (Newbold et. al, forthcoming). The expert panel was not asked to predict changes to fish and shellfish populations under baseline and TMDL conditions separately but rather how the current populations would be affected by the improving water quality. As such, benefits for striped bass, blue crab, and oyster attributes were found by applying the predicted percent change to current stock levels.

The projected change in the number of lakes in the watershed with low algae levels were generated using nutrient loadings taken from the Chesapeake Bay Watershed Model (US EPA, 2010) and applying them to the Northeast SPARROW model (Moore et. al 2011). The result is the number of lakes falling into each of four trophic status categories. In order to present respondents with a single tractable attribute for lake conditions the highest trophic state was

---

[19] In both the baseline and TMDL projections all other socio-economic factors (such as population) are held constant.

[20] The full set of models used by the CBPO to project water quality across the Watershed is documented on the CBPO website (http://www.chesapeakebay.net/about/programs/modeling, accessed June 4, 2015). The WIPs for each jurisdiction can be found at: http://www.epa.gov/reg3wapd/tmdl/ChesapeakeBay/ (accessed June 4, 2015).

described as “high algae levels” and the other three categories as “low algae levels.” The number of lakes falling into the highest trophic category under baseline and TMDL conditions were used to calculate benefits of reducing nutrient loads to lakes.[21]

Given specific changes in outcomes, welfare calculations based on RUM results are well developed in the literature (Hanemann, 1999; Morey, 1999). Following the approach outlined by Holmes and Adamowicz (2003), household WTP is calculated as:

$$WTP^{HH} = \frac{\left(\hat{\boldsymbol{\beta}} ln(\boldsymbol{x^1}) - \hat{\boldsymbol{\beta}} ln(\boldsymbol{x^0})\right)}{-\hat{\gamma}} \tag{6}$$

where $\boldsymbol{x^1}$ and $\boldsymbol{x^0}$ are vectors of the environmental attribute levels projected under the TMDL and baseline scenarios, respectively, and the estimates of $\hat{\boldsymbol{\beta}}$ and $\hat{\gamma}$ come from the RUM regressions estimated in the previous section.[22] In this application $\boldsymbol{x^0}$ corresponds to the status quo attribute levels of the constant baseline version of the survey, and $\boldsymbol{x^1} = \boldsymbol{x^0} + \Delta\boldsymbol{x}$, where $\Delta\boldsymbol{x}$ are the projected improvements presented in Table 14. Using the results of Model 1, the estimated annual household WTP for the improvements projected under the TMDL are presented in the top panel of Table 15.

Similarly, for Model 2, which allows the marginal utility of the environmental attributes to differ across respondents that are users versus non-users of the Chesapeake Bay and Watershed lakes, the household WTP can be calculated as:

---

[21] Thanks to Bryan Milstead of EPA Office of Research Development and Gary Shenk of USGS Northeast Region for their assistance in projecting water quality in Watershed lakes.

[22] Note that the cost attribute is not included in the utility function here since it drops out of the expression when taking the difference between v1 and v0. This is equivalent to assuming zero income effects (Holmes and Adamowicz, 2003).

$$WTP^{HH} = \frac{\hat{\boldsymbol{\beta}}_{\boldsymbol{non}} ln(\boldsymbol{x}^{\mathbf{1}}) + \hat{\boldsymbol{\beta}}_{\boldsymbol{user}} \boldsymbol{ln}(\boldsymbol{x}^{\mathbf{1}}) \cdot \boldsymbol{user} - (\hat{\boldsymbol{\beta}}_{\boldsymbol{non}} \boldsymbol{ln}(\boldsymbol{x}^{\mathbf{0}}) + \hat{\boldsymbol{\beta}}_{\boldsymbol{user}} \boldsymbol{ln}(\boldsymbol{x}^{\mathbf{0}}) \cdot \boldsymbol{user})}{-\gamma} \quad (7)$$

Where $\hat{\boldsymbol{\beta}}_{\boldsymbol{non}}$ is the subvector containing the estimated coefficients on the uninteracted attribute levels, $\hat{\boldsymbol{\beta}}_{\boldsymbol{user}}$ is the subvector containing the estimated coefficients on the user-interacted attribute levels and ***user*** is a vector of dummy variables with the first four elements equal to one if user *i* is a user of the Bay, the fifth element equal to one if user *i* is a user of Watershed lakes, and zero otherwise. The dot product operation in equation (7) indicates element-by-element multiplication.

Table 15. Annual Household WTP for TMDL (2014 dollars)

Model 1	**Model 2**	
	Users	Nonusers
153.6145***	179.7942***	140.2829***
(25.5801)	(32.7696)	(29.1636)
Model 1'	**Model 2'**	
	Users	Nonusers
93.3244***	114.9052***	82.5101***
(28.7002)	(35.6861)	(30.7800)

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

The WTP estimates for all household types are positive and significant. The table also shows that users of both the Bay and Watershed lakes tend to have a higher WTP than nonusers of both these resources. Similar results hold for models 1' and 2' in the lower panel of Table 15, showing point estimates are lower than the corresponding estimates in Models 1 and 2 after controlling for potential biases associated with considerations of the seafood market and lakes outside the Watershed.[23]

To calculate total benefits, WTP is first estimated for a "representative" household (WTP^{RHH}) and then scaled by the population of the study area. Similar to the above exercise, the coefficient estimates from the RUM results are used, but instead of using household-specific values for each variable, the population proportions are entered. This provides an average household WTP, weighted to reflect the distribution of users, income, education, race and ethnicity, in the population of the study area.[24]

For example, instead of the vector of *user* dummy variables in equation (7), the scalar proportions of the population that are users of the Bay and Watershed lakes estimated from the non-response follow up questionnaire enter the equation. Similar calculations are conducted for more complex models that also account for income, education, race, and ethnicity. More formally, household WTP is calculated as:

$$WTP^{RHH} = \frac{(\widehat{\boldsymbol{\beta}}_{\boldsymbol{non}} ln(\boldsymbol{x^1}) + \widehat{\boldsymbol{\beta}}_{\boldsymbol{z}}[ln(\boldsymbol{x^1}) \cdot \boldsymbol{z}] - \{\widehat{\boldsymbol{\beta}}_{\boldsymbol{non}} ln(\boldsymbol{x^0}) + \widehat{\boldsymbol{\beta}}_{\boldsymbol{z}}[ln(\boldsymbol{x^0}) \cdot \boldsymbol{z}]\})}{-\gamma} \quad (8)$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{z}}$ is the coefficient subvector corresponding to the interaction terms between these attributes and the included dummy variables denoting household specific characteristics, which vary by model. The vector $\boldsymbol{z}$ includes the proportions of the population that are, for example, users

[23] Nonlinear Wald tests corresponding to models 2 and 2' fail to reject the null hypothesis that a user and nonuser household have an equal WTP for the TMDL ($\chi^2(1)=0.76$, p=0.3846; and $\chi^2(1)=0.82$, p=0.3661, respectively).
[24] A similar approach is taken by Van Houtven et al. (2014) in their analysis of lake water quality in Virginia.

of the Bay, have a college degree, etc. These population values, shown in Table 16, are plugged into equation (8), accordingly.

Table 16. Population Statistics Used in Estimating WTP of "Representative" Household

Variable	% Population	Source
Bay User	15.3%	Nonresponse follow-up survey
Lake User	14.5%	
Annual Income $0 - $50k	49.4%	US Census 2009 American Communities Survey
Annual Income $50 - $100k	30.3%	
Annual Income >$100k	20.4%	
College Degree or Higher	44.8%	2010 US Census
Black or African American	18.2%	
Hispanic, Latino, or Spanish Origin	12.5%	

The annual WTP for the representative household is calculated using each of the models. Notice that the *SQC* is not included in the total WTP calculations, as this could potentially bias the welfare estimates for the specified attribute improvements. Similarly, in models 1' through 4' the coefficients corresponding to potential considerations of the price and quality of seafood, or lakes outside of the Watershed are not included when calculating WTP.

The representative household's annual WTP estimates are presented in Table 17. The annual WTP estimates for the representative household are fairly robust across the different specifications (Models 1 through 4), ranging from $121 to $157. The WTP estimates are lower under models 1' through 4', ranging from $54 to $94. For both sets of models the household WTP estimates are all statistically significant ($p < 0.01$); the only exception is Model 4' where the corresponding p-value is 0.109. Overall, the results from both sets of models suggest that the complexity of the specifications in accounting for observed household heterogeneity does not materially affect the annual household WTP estimates and there is no clear trend of WTP increasing or decreasing with model complexity.

Table 17. Annual WTP Estimates for the Representative Household (2014 dollars).

Model 1	**Model 2**	**Model 3**	**Model 4**
153.6145***	146.2503***	156.5385***	121.1367***
(25.5801)	(26.4238)	(29.0149)	(31.9006)
Model 1'	**Model 2'**	**Model 3'**	**Model 4'**
93.3244***	87.4356***	94.2640***	54.1739
(28.7002)	(28.6180)	(31.6179)	(33.7703)

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

Estimating Total WTP

To estimate the total economic benefits of the TMDL, the representative household WTP values are multiplied by the corresponding number of households in the population of our study area. Based on the 2010 decennial census conducted by the U.S. Census Bureau, there are a total of 44,353,441 households within the study area. A critical consideration, however, are assumptions regarding the WTP of households that did not respond to the survey. One could posit that some households did not respond to the survey because they do not hold values for improvements to the

Chesapeake Bay and lakes in the Watershed. Such households would have a relatively low WTP, perhaps even zero in the extreme case. On the other hand, if households did not respond for reasons independent of their preferences for improvements in the Chesapeake Bay and Watershed lakes, then the sample of survey respondents (and their estimated WTP) would be representative of the population. For the former case, the representative household WTP estimates are only applied to the proportion of the population equal to the 31% response rate achieved by the main survey. For the latter case, the household WTP estimates are applied to all households in the study frame. Both bounding cases have been implemented in other SP studies calculating total benefits for improvements in water quality and aquatic species (e.g., Van Houtven et al., 2014; Mansfield et al. 2012).

The upper and lower bound total annual WTP estimates are presented for each model in Tables 15 and 16. Across all model specifications and both bounding assumptions on nonresponding households' preferences, the final range of benefits projected to result from the environmental improvements under the TMDL is estimated at $0.745 billion to $6.943 billion per year, where again the lower end, as predicted by Model 4' is statistically insignificant (p=0.109). In one sense this is a conservative range of estimates because it only includes households within the study area, and given the lack of evidence supporting a WTP distance decay gradient, it is likely that households outside the study frame hold a positive WTP for improvements in the Chesapeake Bay.

Table 18. Upper Bound Total Annual WTP Estimates (billions of 2014 dollars): WTP of non-respondents assumed to be similar to respondents

Model 1	Model 2	Model 3	Model 4
6.813***	6.487***	6.943***	5.373***
(1.135)	(1.172)	(1.287)	(1.415)
Model 1'	**Model 2'**	**Model 3'**	**Model 4'**
4.139***	3.878***	4.181***	2.403
(1.273)	(1.269)	(1.402)	(1.498)

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

Table 19. Lower Bound Total Annual WTP (billions of 2014 dollars): WTP of non-respondents assumed to be zero

Model 1	Model 2	Model 3	Model 4
2.127***	2.01894***	2.157***	1.671***
(0.349)	(0.361)	(0.394)	(0.431)
Model 1'	**Model 2'**	**Model 3'**	**Model 4'**
1.283***	1.202***	1.296***	0.745
(0.395)	(0.393)	(0.435)	(0.464)

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

Comparisons of the total WTP estimates in Tables 15 and 16 reveal noticeable differences between the core model results (models 1 through 4) and those from models 1’ through 4’, which eliminate possibly inflated WTP associated with respondents that may have considered omitted variables, namely seafood price/quality and lakes outside the Watershed. The largest factor affecting the final WTP ranges, however, are the assumptions on the WTP of non-responding households. Conditional on these bounding assumptions there is relatively little difference in the total WTP estimates across model specifications.

Since accounting for observed respondent and household heterogeneity makes relatively little difference in the final WTP results, we henceforth focus on Models 2 and 2' for three reasons. First, comparing across all four specifications, Models 2 and 2' always fall within the lower-middle range of the results. Second, these specifications account for a critical dimension of household heterogeneity by allowing preferences to vary across households who use the Chesapeake Bay or lakes in the Watershed for recreation, versus those who are nonusers. Third, although Models 3 and 4 account for additional (and often significant) dimensions of observed household heterogeneity, the numerous interaction terms in these models can be burdensome and risk over-fitting the econometric model. Since such complications do not make much difference in the final WTP estimates, the simpler Model 2 is preferred. Model 2' is included since accounting for the premium associated with respondents who may have considered omitted variables significantly impacts the final WTP estimates. As shown in Tables 15 and 16, the preferred set of models (Model 2 and 2') suggest a total annual WTP of $1.202 to $6.487 billion per year (with 95% confidence intervals ranging from $432 million to $8.784 billion per year).

The total benefits derived from models 2 and 2' are next broken down in terms of the proportion of benefits attributed to improvements in Watershed lakes versus the Bay itself. About half of the total annual benefits estimated from this SP study are attributed to improvements in freshwater lakes in the Watershed (52% and 46% under Models 2 and 2', respectively). Additionally, according to Models 2 and 2', about 80% of the total benefits resulting from the TMDL accrue to households who are nonusers of the corresponding resources.

8. VALIDITY AND SENSITIVITY

Scope tests are a common validity test of SP studies, and assess whether stated responses show that utility is increasing with the quality or quantity of the commodity of interest. The positive and statistically significant coefficients in column 1 of Table 20, our base model, pass an internal scope test: respondents are generally willing to pay more for increases in defined attributes.

Table 20. RUM Regressions Testing for Internal and External Scope: Model 1[a]

	All Choice Questions	First Question Only
Variables	(1)	(2)
ln(clarity)	0.9356**	1.2412
	(0.4586)	(0.8770)
ln(bass)	1.1785**	2.3600**
	(0.5088)	(1.0572)
ln(crab)	2.2465***	2.7316**
	(0.6755)	(1.3913)
ln(oysters)	0.4189**	1.0492**
	(0.1655)	(0.4200)
ln(lakes)	3.9673***	3.1255**
	(0.6350)	(1.5130)
cost	-0.0074***	-0.0112***
	(0.0007)	(0.0020)
status quo (mean)	-1.5335***	-3.5955*
	(0.3551)	(1.9407)
status quo (std dev)	4.2770***	7.4120**
	(0.4705)	(3.2551)
Observations	4,719	1,602
ll	-1.0702e+08	-4.1929e+07

*** p<0.01, ** p<0.05, * p<0.1.

a. Standard errors in parentheses, clustered at the respondent-level. Mixed logit regressions estimated using 200 Halton draws.

To explore external scope, we focus on the first choice occasion thereby removing the potential for anchoring and internally consistent behavior within a respondent. This is not an external scope test in the sense of having two independent subsamples, but if we find the coefficients are significant and of the expected sign when estimated from a cross-section of responses, then the results can be said to pass this less preliminary test of external scope (Carson, 1997). Looking at our model of first responses estimated across respondents in column 2 of Table 20, we find the coefficients are of the expected sign and, with the exception of clarity, are statistically significant.

Another common concern in the literature is that stated choices, and thus the preferences inferred from those choices, may not be consistent across choice occasions. Such inconsistencies can arise from respondents becoming fatigued, learning, or their decision rule may change as they progress through the series of choice questions (Hess et al., 2012; Holmes and Boyle, 2005; Savage and Waldman, 2008). If the parameter estimates are similar across the different choice questions, then this provides some support that the responses are based on stable underlying preferences, which helps validate the overall survey instrument and SP study.

To assess consistency across the choice occasions, a variant of the base model is estimated including interaction terms between the environmental and cost attributes and dummy variables denoting each choice occasion $t = 1, 2,$ or 3. The results are presented in Table 21. There is no clear trend in the marginal utility estimates across the various choice occasions; with slight declines among some attributes and no monotonic trend among others. With the exception of striped bass, statistical tests fail to reject the null hypotheses that the marginal WTP estimates for each attribute are statistically equal across the three choice questions suggesting that respondents' stated decisions are consistent across the choice questions.

Table 21. Marginal WTP: Testing for Consistent Preferences (2014 dollars)[a]

MWTP	Question 1	Question 2	Question 3	H_0: MWTP equal[b]
clarity	37.9148	55.8119*	36.4521	$X^2(2)=0.44$
	(25.5076)	(28.6722)	(27.1725)	p=0.8200
bass	10.5235***	7.2391*	-0.0792	X^2 (2)=4.83*
	(3.2178)	(4.3259)	(3.6884)	p=0.0893
crab	1.4107***	0.9987*	0.6648	X^2 (2)=1.99
	(0.3894)	(0.6070)	(0.4592)	p=0.3706
oysters	0.0286***	0.0092	0.0134	X^2 (2)=3.28
	(0.0076)	(0.0087)	(0.0088)	p=0.1937
lakes	0.1509***	0.1443***	0.1349***	X^2 (2)=0.09
	(0.0447)	(0.0455)	(0.0407)	p=0.9538

*** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses.

[a] Underlying mixed logit regressions estimated using 200 Halton draws, and are specified to account for multiple choice questions answered by each respondent.

[b] Nonlinear Wald test of the null hypothesis that the MWTP estimates inferred from each individual question are equal.

9. CONCLUSION

We present the results of a stated preference survey to examine values for programs to address excess nutrient and sediment loadings in the Chesapeake Bay for households in 17 eastern states and the District of Columbia. These estimates are linked to a suite of hydrological and ecological models that project how these five environmental attributes may change over time under the TMDL and baseline conditions. This allows us to estimate the total economic benefit of the management practices required to meet the TMDL. Understanding total public values for ecosystem resources, including the more difficult to estimate non-use values, is necessary to determine the full range of benefits associated with reductions in nutrient and sediment loading.

Results from our preferred specifications (Models 2 and 2') show that a representative household in the study area is willing to pay between $87 and $146 per year for water quality improvements from the TMDL. As expected, individual users are willing to pay more for improvements than individual non-users. Nonetheless the WTP of non-users in aggregate, constitutes a large proportion of the total benefits. An estimated 71% of the households in the study area are non-users of the Chesapeake Bay or Watershed lakes. The survey data used in the main analysis pass tests of scope and preference consistency.

Applying the representative household WTP estimates from our preferred models to the population of the study area yields a range of total WTP from about $1.202 to $6.487 billion per year, depending upon specification of the econometric model and what is assumed about WTP of non-responders. About 46% to 52% of the benefits stem from improvements in freshwater lakes within the Watershed. This is a key finding – the costs of the TMDL accrue to states throughout the Watershed, not just those in the immediate vicinity of the Bay, and our results show the benefits of the TMDL are not limited to the Bay itself. Further, about 80% of the total benefits accrue to nonusers of the Bay and freshwater lakes in the Watershed. Together these results suggest that non-use values and ancillary benefits are key components of the total benefits of policies to improve water quality in the Chesapeake Bay and Watershed. Such findings emphasize the need for comprehensive benefits analyses, including well-constructed SP studies, when assessing the social benefits generated by actions to improve large, interconnected, and iconic natural resources, like the Chesapeake Bay.

References

Banzhaf, H. Spencer, Dallas Burtraw, David Evans, Alan Krupnick (2006). Valuation of natural resource improvements in the Adirondacks. *Land Economics* 82: 445-464.

Bockstael, Nancy E., Kevin E. McConnell, Ivar E. Strand (1988). *Benefits from Improvements in Chesapeake Bay Water Quality, Volume III*. Washington, DC: U.S. Environmental Protection Agency.

Bockstael, Nancy E., Kevin E. McConnell, Ivar E. Strand (1989). Measuring the benefits of improvements in water quality: The Chesapeake Bay. *Marine Resource Economics* 6: 1-18.

Boyd, James, Alan Krupnick (2009). The definition and choice of environmental commodities for nonmarket valuation. *RFF Discussion Paper* 09-35, Resources for the Future.

Boyd, James, Alan Krupnick (2013). Using ecological production theory to define and select environmental commodities for nonmarket valuation. *Agricultural and Resource Economic Review* 42(1): 1-32.

Brouwer, Roy (2008). The Potential Role of Stated Preference Methods in the Water Framework Directive to Assess Disproportionate Costs. *Journal of Environmental Planning and Management* 51(5): 597-614.

Carson, Richard T. (1997) Contingent valuation surveys and tests of insensitivity to scope. In *Determining the Value of Non-market Goods: Economics, Psychology, and Policy Relevant Aspects of Contingent Valuation.* Edited by R.J. Kopp, W. Pommerhene, and N. Schwartz. Boston, Kluwer, pp127-163.

Cropper, Maureen, William Isaac (2011). The benefits of achieving the Chesapeake Bay TMDLs (Total Maximum Daily Loads): A scoping study. Washington, D.C.: Resources for the Future.

Cummings, Ronald G., P. Ganderton, and T. McGuckin. (1994). Substitution Effects in CVM Values. *American Journal of Agricultural Economics* 76: 205-214.

Cummings, Ronald G., and Laura O. Taylor. (1999). Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *American Economic Review* 89(3): 649-665.

Dillman, D.A. (2008). *Mail and internet surveys: The tailored design method*. New York: John Wiley and Sons.

Greene, William H. (2003). *Econometric Analysis*. New Jersey: Prentice-Hall.

Hanemann, W.M. (1999). Welfare Analysis with Discrete Choice Models. In *Valuing Recreation and the Environment in Theory and Practice*. Edited by C. Kling and J. Herriges. Northampton, MA: Edward Elgar.

Herriges, Joseph, Catherine Kling, Chih-Chen Liu, Justin Tobias (2010). What are the consequences of consequentiality? *Journal of Environmental Economics and Management,* 59: 67-81.

Hess, S., D. Hensher, and A.J. Daly. (2012). Not Bored Yet – Revisiting Respondent Fatigue in Stated Choice Experiments. *Transportation Research Part A* 46(3): 622-644.

Hicks, Robert, James E. Kirkley, Kevin E. McConnell, Winifred Ryan, Tara L. Scott, and Ivar Strand (2008). *Assessing stakeholder preferences for Chesapeake Bay restoration options: A stated preference discrete choice-based assessment* (pp. 1-56). Annapolis, MD: NOAA Chesapeake Bay Office, National Marine Fisheries Service and Virginia Institute of Marine Science.

Hogg, Robert V., Joseph W. McKean, and Allen T. Craig (2005), Introduction to Mathematical Statistics, 6th edition, Upper Saddle River, NJ: Pearson Prentice Hall.

Hole, Arne Risa. (2007). Fitting Mixed Logit Models by Using Simulated Maximum Liklihood. *The Stata Journal* 7(3): 388-401.

Holmes, Thomas P., and Kevin J. Boyle (2005). Dynamic Learning and Context-Dependence in Sequential, Attribute-Based, Stated-Preference Valuation Questions. *Land Economics* 81(1): 112-126.

Holmes, Thomas P., and Wiktor L. Adamowicz (2003). Attribute-Based Methods. In *A Primer on Nonmarket Valuation*, Edited by Patricia A. Champ, Kevin J. Boyle, and Thomas C. Brown, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Johnston, Robert J, Thomas A Grigalunas, James J Opaluch, Marisa Mazzotta, Jerry Diamantedes (2002). Valuing estuarine resource services using economic and ecological models: The Peconic Estuary System study. *Coastal Management* 30: 47–65.

Johnston, Robert J., Daniel Jarvis, Kristy Wallmo, Daniel K. Lew (2015). Multi-scale Spatial Pattern In Nonuse Willingness to Pay: Applications to Threatened and Endangered Marine Species. *Land Economics* 91(4): in press.

Johnston, Robert J, James J Opaluch, Marisa J Mazzotta, Gisele Magnusson (2005). Who are resource non-users and what can they tell us about non-use values? Decomposing user and non-user willingness to pay for coastal wetland restoration. *Water Resources Research,* 41(7).

Johnston, Robert J., and Mahesh Ramachandran (2014). Modeling Spatial Patchiness and Hot Spots in Stated Preference Willingness to Pay. *Environmental and Resource Economics*, 59: 363-387.

Johnston, Robert J., Eric T. Schultz, Kathleen Segerson, Elena Y. Besedin, and Mahesh Ramachandran (2013). Stated Preferences for Intermediate versus Final Ecosystem Services: Disentangling Willingness to Pay for Omitted Outcomes. *Agricultural and Resource Economics Review* 42(1): 98-118.

Johnston, Robert J., Stephen K. Swallow, Timothy J. Tyrrell, and Dana M. Bauer. (2003). Rural Amenity Values and Length of Residency. *American Journal of Agricultural Economics* 85(4): 1000-1015.

Kragt, Marit E, Jeffrey W Bennett (2011). Using choice experiments to value catchment and estuary health in Tasmania with individual preference heterogeneity. *The Australian Journal of Agricultural and Resource Economics* 55: 159-179.

Lipton, Douglas (2004). The value of improved water quality to Chesapeake Bay boaters. *Marine Resource Economics* 19*:* 265-270.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Economics.* New York: Cambridge University Press.

Mansfield, C.**,** Van Houtven, G.**,** Hendershott, A.**,** Chen, P.**,** Porter, J.**,** Nourani, V.**, &** Kilambi , V. (2012). Klamath River Basin restoration: Nonuse value survey. Final report: Prepared for the U.S. Bureau of Reclamation. Research Triangle Park, NC: RTI International.

Metcalf, Paul J., William Baker, Kevin Andrews, Giles Atkinson, Ian J. Bateman, Sarah Butler, Richard T. Carson, Jo East, Yves Gueron, Rob Sheldon, and Kenneth Train. (2012). An Assessment of the Nonmarket Benefits of the Water Framework Directive for Households in England and Wales. *Water Resources Research* 48(3).

Moore, R., Johnston, C., Smith, R., & Milstead, B. (2011). Source and delivery of nutrients to receiving waters in the northeastern and mid-Atlantic regions of the United States. *Journal of the American Water Resources Association 47*(5): 965-990.

Newbold, Steve C., Massey, D. Matthew, Moore, Chris M., Commercial and Recreational Fishing Benefits of the Chesapeake Bay Total Maximum Daily Loads, *In Preparation.*

Phaneuf, Daniel J, Roger H von Haefen, Carol Mansfield, George Van Houtven (2013). Measuring nutrient reduction benefits for policy analysis using linked non-market valuation and environmental assessment models, Final Report on Stated Preference Surveys. Report to the US EPA.

Roberts, David C., Tracy A Boyer, Jason I. Lusk (2008). Preferences for environmental quality under uncertainty. *Ecological Economics* 66: 584-593.

Rolfe, John, and Jill Windle (2012). Distance Decay Functions for Iconic Assets: Assessing National Values to Protect the Health of the Great Barrier Reef in Australia. *Environmental and Resource Economics* 53: 347-365.

Savage, Scott J., and Donald M. Walman (2008). Learning and Fatigue During Choice Experiments: A Comparison of Online and Mail Survey Modes. *Journal of Applied Econometrics* 23: 351-371.

Train, Kenneth E. (2009). *Discrete Choice Methods with Simulation*. New York: Cambridge University Press.

Van Houtven, George, Carol Mansfield, Daniel J. Phaneuf, Roger von Haefen, Bryan Milstead, Melissa A. Kenny, Kenneth H. Reckhow (2014). Combining Expert Elicitation and Stated Preference Methods to Value Ecosystem Services from Improved Lake Water Quality. *Ecological Economics* 99: 40-52.

Van Houtven, George, John Powers, Subrendu Pattanayak (2007). Valuing water quality improvements in the United States using meta-analysis: Is the glass half-full or half-empty for national policy analysis? *Resource and Energy Economics* 29: 206-228.

Viscusi, W Kip, Joel Huber, Jason Bell (2008). The economic value of water quality. *Environmental and Resource Economics* 41: 169-187.

Windle, Jill, John Rolfe (2004). Assessing values for estuary protection with choice modelling using different payment mechanisms, valuing floodplain development in the Fitzroy Basin, Research Report No.10. Faculty of Business and Law, Central Queensland University, Emerald, Queensland.