Chesapeake Bay Watershed Stated Preference Study

Draft Report

US Environmental Protection Agency
National Center for Environmental Economics

October 27, 2014

Contact: Chris Moore
moore.chris@epa.gov

Contents

List of Tables

List of Figures

1. Introduction

The Chesapeake Bay watershed encompasses 64,000 square miles in parts of six states and the District of Columbia. It is the largest estuary in the United States and the third largest in the world. 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. Significant progress has been made over that period; however, a pollution budget, called a Total Maximum Daily Load (TMDL), is necessary to continue progress toward the goal of a healthy Bay.

The US Environmental Protection Agency (EPA) desires a monetary measure of the benefits associated with environmental improvements expected from recent TMDL requirements for the Chesapeake Bay promulgated in December 2010. The TMDL establishes limits on nutrient inflows into the Chesapeake Bay from the surrounding Chesapeake Bay Watershed. The District of Columbia and the six states within the Chesapeake Bay Watershed submitted Watershed Implementation Plans describing the actions each state intends to implement by 2025 to meet TMDL limits.

Benefits from meeting the TMDL for the Chesapeake Bay will accrue to those who live near the Bay or visit for recreation, those who live near or visit lakes and rivers in the watershed, and those who live further away and/or may never visit the Bay but have a concern for these water bodies. While benefits from the first two categories can be measured using hedonic property value, recreational demand, and other revealed preference approaches, only stated preference methods can capture non-use benefits (i.e., benefits to those who may never visit the Bay or affected lakes and rivers). Understanding total public values for ecosystem resources in the Bay, 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. Because non-use values may be substantial, failure to estimate such values may lead to improper inferences regarding these benefits and associated costs.

Transferring estimates from other studies based in other estuaries (i.e., benefit transfer) is not advised in this case as these results are unlikely to accurately or completely capture willingness to pay for TMDL-related improvements in the Chesapeake Bay Watershed given the unique character of this water resource and the goods and services it provides. Further, there are few stated preference studies in the published literature focusing on the Chesapeake Bay, and no studies specifically addressing the environmental improvements predicted under the TMDL.

To more completely capture the total benefits of the TMDL, including the value to both users and non-users of the Chesapeake Bay, EPA conducted a stated preference survey in the summer of 2014 of residents living in 17 east coast states and the District of Columbia. Seven of the states included are located within the Chesapeake Bay Watershed. The Chesapeake Bay Stated Preference Survey (OMB control number 2010-0043) was designed and administered by mail with contractual support provided by Abt Associates, Inc. EPA designed the survey with multiple objectives in mind including understanding how individuals value improvements to specific attributes in the Chesapeake Bay and its Watershed lakes (e.g., water clarity; adult populations of striped bass, blue crab, oysters, and lake conditions); understanding how the above values depend on the future baseline level of water quality in the Chesapeake Bay and its Watershed; and understanding how values vary with respect to individuals' attitudes, awareness, and demographic characteristics.

This report presents the results from the Chesapeake Bay Stated Preference survey. Section 2 provides additional background on stated preference methods and reviews related studies. The survey design is described in section 3, including the final survey instrument and attributes, and the survey development process. Section 4 details the sample frame, stratification, and approach to data collection, while section 5 provides a summary of the data. Analysis of choice experiment data is defined and described in section 6. The survey results are presented in section 7 and includes extrapolation of household willingness to pay to the relevant population in order to obtain a total benefits estimate. Section 8 provides information on validity tests and sensitivity analysis of key assumptions. Finally, a summary and discussion is found in section 9.

2. Theory and Background

To estimate the benefits of an environmental improvement, one needs to know the economic value (or total dollar value) an individual would voluntarily pay for the improvement. Because environmental improvements are generally not traded in the marketplace, economists often must rely on non-market valuation techniques to derive these *willingness to pay* (WTP) measures. The stated preference method relies on data from surveys that ask respondents to consider various hypothetical scenarios and state their preferences directly. Provided that respondents answer questions truthfully, stated preference data can then be used to estimate WTP for the good in question. Unlike their revealed preference counterparts which infer WTP estimates from observed behavior in related markets, stated preference techniques can be tailored to match specific policy scenarios and can be used to estimate both use and non-use values.

In spite of its flexibility, the stated preference approach has been criticized due in large part to the hypothetical nature of the decisions respondents are asked to make and the potential for related systematic biases (e.g., hypothetical bias and non-response bias). However, these biases can be mitigated through careful survey development and design, including thorough definitions of the commodity and payment vehicle. Regardless, stated preference results should be subject to reliability and validity tests prior to their application to policy decisions. While the methodology has its critics (e.g. Hausman 2012), stated preference surveys have been applied to a wide array of environmental goods and there is general acceptance of the method within academic circles and for public policy analysis (Carson 2012, Kling et al. 2012).

2.1 Choice Experiments

Stated preference is an umbrella term used to describe survey methods for eliciting WTP values for non-market goods. There has been considerable confusion in terminology applied to various stated preference techniques (Louvier and Carson 2011), however a distinction is generally made between contingent valuation and choice experiments (sometimes known as discrete choice experiments.) Contingent valuation (CV) typically consists of a single binary choice question, perhaps with a follow up that is conditional on the first response (e.g., double-bounded), that estimates willingness to pay for a given attribute or set of attributes taken as a whole. Often the dichotomous choice question is framed as a referendum. Choice experiments (CE), by contrast, rely upon a sequence of multinomial choices to estimate implicit prices for specific attributes.

CEs have been increasingly used to estimate the value of changes in environmental quality (Alpizar et al. 2001; Bennett and Adamowicz 2001; Carson and Czajkowski 2014). Respondents in a CE are presented with a series of questions describing the outcome of alternative policy scenarios. These choice sets are characterized by a set of attributes with levels that vary across these scenarios and represent available choice sets. In order to estimate willingness to pay, one of these attributes is a cost to the respondent. Respondents are asked to choose their preferred option from the available choices and are expected to make tradeoffs among the levels of the attributes, including costs. By evaluating the tradeoffs made by respondents one can infer their relative value and, by using the cost attribute, estimate marginal willingness to pay for a change in each attribute.

Choice experiments can gather more information on preferences per respondent than can a contingent valuation survey by varying attribute levels across multiple choice questions. CE may also be more advantageous for benefit transfer, particularly for policies that might address only a subset of the attributes included in the original study. CV provides willingness to pay for an aggregate of several factors, but these factors are not valued separately. By contrast, the vector of implicit prices estimated in CE represent marginal willingness to pay values for each attribute that can be applied separately to evaluate a range of policy outcomes. This is subject, of course, to the usual caveats and best practices associated with benefit transfer more broadly (e.g., Bergstrom and Taylor 2006; Johnston and Rosenberger 2010).

2.2 SP Applications to Water Quality

There are a wide variety of SP applications to water quality, most often focusing on specific bodies of water or locations. For reviews of the literature see Johnston, et al. 2005 and Van Houtven, et al. 2008. One key consideration in the design of SP studies of water quality is the characterization of the water quality itself. Many studies use indicators of achieving water quality criteria for designated uses, such as fishing, swimming, or irrigation, that depend upon pollutant levels (e.g.,Lindsey, Paterson, and Luger 1995; Smith, Desvousges, and Fisher 1986; Viscusi, Huber, and Bell 2008). Other studies combine several variables, such as pH and turbidity, into an aggregate index that may provide a more complete characterization of water quality (e.g., Jeon et al. 2005, Desvousges, Smith, and Fisher 1987, and Carson and Mitchell 1993). Water quality indices may facilitate benefit transfer (Griffiths, et al. (2012)) and can be used in meta-analyses (e.g., Van Houtven, Powers, and Pattanayak, 2007). Benefits analysis, however, can be sensitive to the choice of index (Walsh and Wheeler, 2013.)

Another approach to water quality valuation in stated preference is to specify ecosystem service endpoints as attributes affected by policy. These include physical characteristics such as water clarity, color, and odor, as well as species abundance and impacts on threatened or endangered species. This approach has the advantage, perhaps, of being most salient and recognizable for survey respondents, but such measures need to be linked to biophysical measures of water quality (Van Houtven, et al. 2014). While there are no standard protocols for defining the effects of ecological changes in stated preference surveys, recent analyses distinguish between "inputs" (environmental features or conditions that, via natural processes, are converted into different features) and "endpoints," or biophysical outputs that are more "directly and economically meaningful" to firms and households (Boyd and Krupnick, 2009, 2013). Generally, endpoints are more appropriate as attributes for stated preference studies because use of inputs leaves the respondent's ecological production function, i.e., how inputs are converted into endpoints, implicit and unobservable.

In all cases for stated preferences studies it is vital that attributes are defined in a way that is salient to respondents and is expressed in metrics that respondents understand and can use to make tradeoffs between attributes. This metric must also be quantitatively linked to endpoints or attributes that respondents care about directly (i.e., enters the utility function.). Finally, for application in benefits analysis, changes in the attributes must be derivable from physical models of the environment affected by policy alternatives. The survey used in this study is based on ecological endpoints. Changes in the Chesapeake Bay that result from actions taken under the TMDLs are characterized by expected changes in the populations of striped bass, blue crabs, and oysters, as well as changes in the clarity of Chesapeake Bay waters. Water quality changes in lakes and reservoirs in the Chesapeake Bay Watershed are characterized by average algae levels, with lakes being classified as either "low" or "high" algae lakes. These attributes were found to be salient and meaningful to respondents in focus groups and interviews, and can be estimated from available models.

2.3 SP Applications for estuaries and the Chesapeake Bay

There are many 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. However, there are relatively few applications of choice experiment methods to estuaries. 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 dollar values for an additional acre of farmland, undeveloped land, wetlands, shellfishing areas, and aquatic grasses. Kragt and Bennett (2011) employ a choice experiment to estimate community preferences for national resource management options in the estuary of Georges Bay, Tasmania, using "area of healthy seagrass beds" as a proxy for estuary condition. Windle and Rolf (2005) estimate non-use values associated with protecting environmental quality in the Fitzroy estuary in the Great Barrier Reef catchment area. The study intentionally sampled respondents distant from the estuary in order to focus on estimating values to non-users.

The Chesapeake Bay, the largest estuary in the US, is an iconic resource with a long history of restoration efforts and is relatively well studied. However, no study or set of studies provides a comprehensive estimate of non-use values or values associated with specific improvements in the Chesapeake Bay and the Chesapeake Bay Watershed, particularly those likely to result from the TMDLs or other policies addressing nutrient pollution.

Cropper and Isaac (2011) provide an extensive review of valuation studies relevant to improving water quality throughout the Chesapeake Bay. Very few studies attempt 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. There is no clear way to link "swimmability" to the improvements expected from the TMDLs, making it difficult to use this study for valuation. Further, the study sample was limited to the Washington-Baltimore area. Lipton et al. (2004) estimate willingness to pay of non-users for restoring oyster reefs in the Chesapeake Bay using a broad, but non-random, sample that includes most mid-Atlantic states. A choice experiment by 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 are possible.

While not directly comparable to our study, a number of studies use other revealed preference approaches to value water quality changes in the Chesapeake Bay, including hedonic property methods (Leggett and Bockstael, 2000; Poor et al., 2007; and Van Houtven, 2009) and travel cost methods (Bockstael et al., 1988, 1989; Lipton and Hicks, 1999; Hicks and Strand, 2000; Lipton and Hicks, 2003; Lipton, 2004; Massey et al., 2006). Other studies use benefit transfer to estimate recreation benefits of improved water quality in the Chesapeake Bay (e.g., Krupnick, 1988; Morgan and Owens, 2001; Van Houtven, 2009). These studies collectively address benefits from recreational fishing, swimming or beach visits, and boating, but this is a small subset of the kinds of benefits expected from broad requirement such as those found in the TMDL.

Several recent studies have applied stated preference methods to value environmental quality in lakes. Banzhaf et al. (2006) use contingent valuation to estimate willingness to pay to reduce acidification at a collection of lakes in the Adirondacks. A referendum style CV question was also used in 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 choice experiment for a single lake using attributes of algae bloom status and water level, with the attributes presented both probabilistically and with certainty. Viscusi, et al. (2008) include a choice experiment 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).

Phaneuf, et al. (2013) employ both a CV referendum question and a choice experiment approach to value improvements in lake water quality within respondents' home state. Water quality was characterized by five trophic categories characterized by water color, clarity, odor, type and abundance of fish, and size and frequency of algae blooms. The choice experiment questions were provided to a subset of users, respondents who had visited a lake recently or planned to do so in the near future. In contrast, our own study includes lake water quality as one attribute characterized by the percentage with high versus low algae levels, although the descriptions include some of the attributes in Phaneuf, et al. (e.g., water color, clarity).

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. Van Houtven et al. also provide a case study showing how their results could be applied to lake water quality improvements in the Virginia area expected to result from requirements established in the 2010 TMDL for the Chesapeake Bay. The case study relies upon expert elicitation to translate changes in pollutant concentrations to changes in the percentage of lakes in each of the trophic categories described to respondents.

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 Smith and Milstead 2011). This study fills a key gap in the literature by valuing a comprehensive set of water quality outcomes for the Chesapeake Bay Watershed, including lakes, the Chesapeake Bay itself, and the tidal sections of its tributaries. Such an approach is essential to evaluating the benefits of policies such as TMDLs that have broad effects across all of these 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; values for non-users can be considered a lower bound on non-use value for these improvements.

The design of this study allows for consideration of how willingness to pay for water quality improvements is affected by baseline trends in the absence of new policies. While we rely upon a constant (i.e., no change from present state) baseline for our primary results we also report willingness to pay for declining and improving baseline versions of the survey. Further, our sampling frame allows us to consider how the value of water quality improvements varies between households within the Chesapeake Bay Watershed and those outside of it, including those in states that contain no part of the watershed. Finally, the attributes and metrics used in our survey instrument can be linked directly to estimated physical changes predicted by water quality and other models. This is essential for applied policy analysis.

3. Survey Instrument Design and Development

The survey design process began with a kick-off workshop held at Resources for the Future on October 31 and November 1, 2011, during which non-market valuation experts provided feedback and advice on conducting a study to value the anticipated improvements. The EPA worked with expert consultants throughout the design process, conducted multiple focus groups and one-on-one interviews and had the survey instrument and study design peer reviewed by three external experts. Finally, a pilot study was conducted in advance of the full scale administration after which only very minor changes were made to the survey instrument.

The survey instrument and study design were also subject to review by the Office of Management and Budget (OMB) under the Paperwork Reduction Act (PRA). As part of the review process, the EPA's information collection request (ICR) was subject to three public comment periods (77 FR 43822, 78 FR 9045, 78 FR 38713). OMB approved the pilot study based on EPA's ICR and response to public comments and approved the full scale administration of the survey after reviewing the results of the pilot.

3.1 Survey Objective and Key Features

The objective of the stated preference study is to estimate WTP for nutrient and sediment loading reductions in the Chesapeake Bay Watershed including use and non-use values. The population of interest includes not only households within the Chesapeake Bay Watershed, but given the iconic nature and ecological importance of the Chesapeake Bay, the WTP of households in all States within 100 miles of the east coast of the U.S.

The choice experiment questions presented status quo and policy scenarios that were represented by a set of 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 a result of the loading reductions. Four attributes relate to the Bay and one attribute relates to freshwater lakes in the watershed. An additional attribute describes annual household costs. The attributes are:

- Water clarity in the Chesapeake Bay in feet of visibility
- Striped bass population in the Chesapeake Bay
- Blue crab population in the Chesapeake Bay
- Oyster population in the Chesapeake Bay
- Number of lakes in the Chesapeake Bay Watershed with low algae levels
- Permanent increase in the cost of living starting next year, shown per month and per year

The levels of the environmental attributes were chosen based on results of the Chesapeake Bay Watershed Model (US EPA 2010) and an internal effort to model fish and shellfish populations (see chapter *#TBD* for a description) in the case of the Bay attributes and the Northeast Lakes Model (Moore Smith and Milstead 2011) in the case of the lakes attribute. Cost levels were chosen to ensure adequate coverage of the WTP distribution without truncating from above and were based on focus group and pilot study results.

3.2 Survey Instrument

Informed respondents tend to provide more reliable answers to SP surveys (Boyle 1989; Bergstrom Stoll and Randal 1990) so it has become standard practice to include an information section describing the affected resources, the changes being considered, and how those changes would affect environmental conditions. Keeping the information section as brief as possible reduces respondent fatigue and improves response rates. Presenting the information in an objective way that educates the respondent without influencing their responses to the WTP questions is imperative to generating credible WTP estimates. The survey instrument includes two pages of information about the Chesapeake Bay, the surrounding watershed, the effect of nutrient and sediment pollution, and questions to reinforce the material.

Following the information section respondents are introduced to the attributes that appear in the choice questions and given their current levels and baseline conditions according to the version of the survey they received. The survey then describes the practices that would improve conditions and how the costs of these practices would be passed onto households. This section includes text noting several factors that the practices would not affect, in order to deter respondents from considering attributes not included in the choice questions. 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.

Based on expert consultation and focus groups the survey includes three choice questions. Figure 1 provides an example from the constant baseline version. 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 one of the three options.

7. Please vote for one of the three options below. *(<u>Mark one box</u> at the bottom to indicate which option you would prefer.)*

	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)	**24** million fish (no change)	**30** million fish (25% increase)
Blue Crab Adult Population	**250** million crabs (no change)	**285** million crabs (14% increase)	**328** million crabs (31% increase)
Oysters Population	**3,300** tons (no change)	**3,300** tons (no change)	**3,300** tons (no change)
Watershed Lakes Lakes with <u>low</u> algae levels	**2,900** lakes (no change)	**3,850** lakes (33% increase)	**3,850** lakes (33% increase)
Your Cost of Living Permanent cost increase for your household starting next year	**$0 every year**	**$60 every year** or $5.00 every month	**$180 every year** or $15.00 every month
Your Vote Please mark <u>one</u> of the boxes to the right	Option A ☐	Option B ☐	Option C ☐

Figure 1 Example Choice Question

Following the choice questions is a set of debriefing questions. These questions are used 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. All debriefing questions use a Likert scale to identify:

- Protest responses – respondents vote for status quo because they reject the provision mechanism, payment vehicle, or some other aspect of the choice scenario other than the policy outcome and costs
- Scenario rejection – the respondent did not accept some aspect of the choice scenario when answering the question (e.g. did not answer as if they would actually have to pay or did not answer as if the attribute improvements would actually be achieved)
- Symbolic responses – respondents vote to show general support or disapproval for the motivation of the policy (e.g. "warm glow") rather than the specific improvements shown

Following the debriefing questions is a set of questions that asks about participation in outdoor recreation on the Chesapeake Bay. Finally, the survey collects demographic data from the in order to compare the sample with the population and to provide conditioning variables for WTP estimation.

3.3 Experimental Design

The purpose of the experimental design was to collect data sufficient to estimate an independent WTP for each survey cell (sampling strata-baseline version combination). In each choice question options B and C, the alternatives to the status quo baseline (option A), are characterized by different levels for the six attributes, including cost. The experimental design is characterized by three levels for each of the five environmental attributes and six levels for costs [$3^5 \times 6$] for each alternative option, or [$3^{10} \times 6^2$] for Options B and C together.

To construct a preliminary main effects design with 72 profiles that is sufficiently flexible to estimate alternative specific main effects and response patterns (i.e., a non-generic design), we begin with a 3^5x6 orthogonal fractional factorial design with 144 profiles. We then combined the elements of this design into pairs that would reflect trade-offs at the margin (i.e., improvements in the attributes that are attained at the cost of decrease in other environmental attributes and/or increase of the overall cost of the program). Finally, these pairs were blocked in such a way that variability of the environmental and cost attributes within a block would be maximized (and hence the main effects would not be confounded with the block effects). The result is a design with 72 profiles.

Three choice questions per survey allows the 72 profiles to be included (orthogonally blocked) in 24 unique survey booklets. The attribute levels applied within surveys are summarized in table 1.

Table 1 Attribute Levels Included in Each Survey Version

Attribute	**Attribute Levels**						
	Baseline	**1**	**2**	**3**	**4**	**5**	**6**
	Declining Baseline						
Water Clarity (feet)	2	3	3.5	4.5	-	-	-
Adult Striped Bass (millions)	21	24	30	36	-	-	-
Adult Blue Crab (millions)	235	250	285	328	-	-	-
Oysters (tons)	2,800	3,300	5,500	10,000	-	-	-
Low Algae Level Lakes	2,300	2,900	3,300	3,850	-	-	-
Annual Household Cost	-	$20	$40	$60	$180	$250	$500
	Constant Baseline						
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	-	$20	$40	$60	$180	$250	$500
	Improving Baseline						
Water Clarity (feet)	3.3	3.3	3.5	4.5	-	-	-
Adult Striped Bass (millions)	26	26	30	35	-	-	-
Adult Blue Crab (millions)	260	260	312	340	-	-	-
Oysters (tons)	4,300	4,300	5,500	10,000	-	-	-
Low Algae Level Lakes	3,100	3,350	3,600	3,850	-	-	-
Annual Household Cost	-	$20	$40	$60	$180	$250	$500

3.4 Survey Development

EPA consulted with several experts in SP and non-market valuation to develop the survey instrument and study design. Alan Krupnick (Resources for the Future), Maureen Cropper (Resources for the

Future, University of Maryland), and Elena Besedin (Abt Associates) have provided guidance and feedback throughout the development of the survey.

As noted earlier, the survey design phase of this project began with a workshop co-sponsored by the EPA on October 31 and November 1, 2011 at Resources for the Future. The main purpose of the workshop was to gather scholars who are working on estimating the costs and benefits of water quality improvements in the Bay to exchange ideas. Although the workshop topics were broader than stated preference techniques, the agenda did include presentations on the estimation of benefits using stated preference methods. Participants included academics from major universities within the Chesapeake Bay Watershed (MD, VA, DE, WV, PA); representatives from several NGOs including the Chesapeake Bay Foundation and the Chesapeake Bay Trust; as well as participants from several government agencies (EPA, NOAA, DOI, and USDA).

As part of the planning and design process for this collection, EPA conducted a series of 10 focus groups and 72 cognitive interviews. Eight of the ten focus groups were in venues located inside the Chesapeake Bay Watershed, in locations close to Bay itself and further upstream. Two other focus groups were conducted in North Carolina, a state that is entirely outside of the Chesapeake Bay Watershed. Of the cognitive interviews, 57 were conducted at locations in the watershed, and 15 were conducted in parts of Pennsylvania that were outside the Chesapeake Bay Watershed. While early focus group sessions were used to narrow the list of possible attributes in a survey and the kinds of information respondents would need to answer the questions, later sessions and cognitive interviews were employed to test the draft survey materials. These consultations with potential respondents were critical in identifying sections of the questionnaire that were redundant or lacked clarity and in producing a survey instrument that is objective and meaningful to respondents.

The survey instrument has undergone peer review by three leading scholars specializing in stated preference surveys for estimating benefits associated with environmental improvements: Kevin Boyle, Professor, Department of Agricultural and Applied Economics, Virginia Tech University; John Whitehead, Professor and Chair, Department of Economics, Walker College of Business, Appalachian State University; and Robert Johnston, Director, George Perkins Marsh Institute, Professor, Department of Economics, Clark University.

EPA conducted a pretest of the survey in November and December 2013 in which 900 surveys were sent out, 300 to each of three geographic strata (see section 4 for a description of the sample stratification). The overall response rate was 34%: 38% in the Bay States, 33% in the Watershed States, and 30% in the Other Eastern States. Pretest results were favorable, indicating sensitivity to scope, both internal and external, and several indications of theoretical validity. Only one very minor change was made after the pretest, replacing one early question to reinforce understanding of the lakes attribute.

3.5 Non-response Bias Study

In order to ascertain if and how respondents and non-respondents differ EPA conducted a non-response bias study 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.

A subset of the questions from the main questionnaire was selected for the non-response bias study survey, as 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.

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.

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

4. Sample and Data Collection

EPA administered the survey 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 East coast U.S. states that include 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, as indicated in table 2 and figure 2.

Table 2 Sample Stratification

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

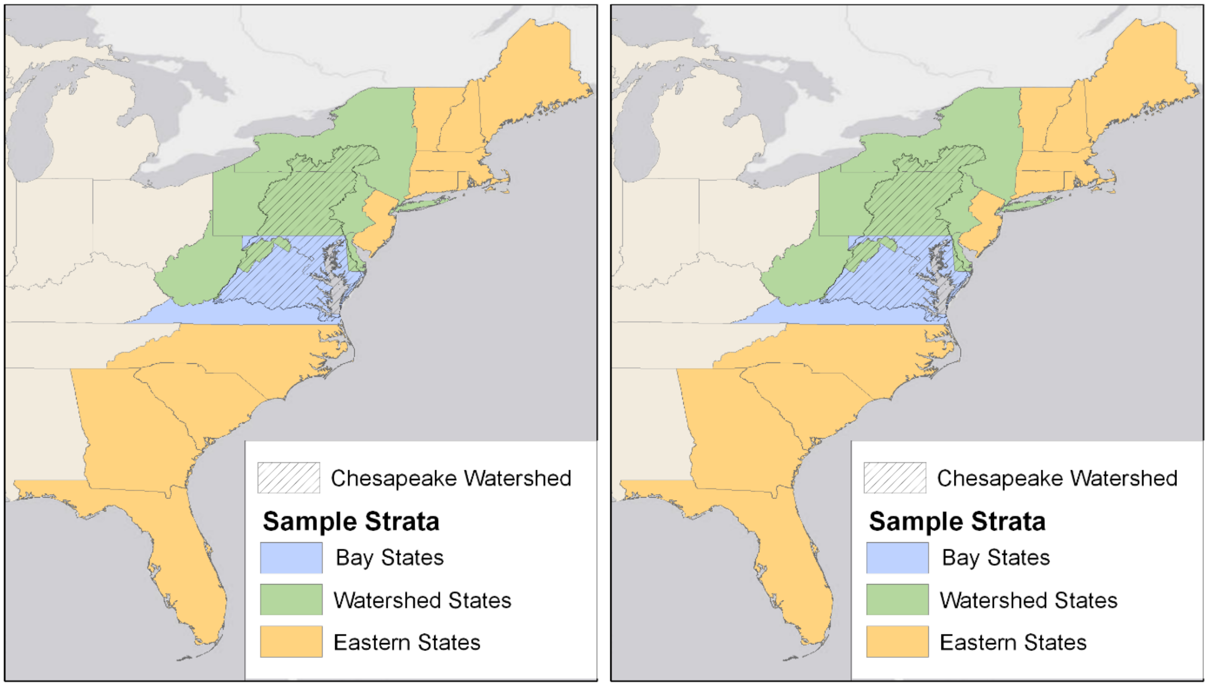

Figure 2 Sampling Frame and Strata

The states of the Bay and Watershed strata were chosen based on their immediate proximity to the Bay and/or lakes, streams and rivers in its watershed. Households in these areas are more likely to hold use values for improvements to the Chesapeake Bay and its watershed than those farther away. The Other East Coast states lie within 100 miles of the Atlantic Ocean. Residents of these states are more likely to be familiar with estuarine issues. At the same time, the greater distance between these states and the Chesapeake Bay will improve the survey's ability to capture values for non-users and to test how quickly the values decrease or "decay" with distance from the Bay region.

The sampling frame is the United States Postal Service Computerized Delivery Sequence File (DSF), the standard frame for address-based sampling (Iannacchione, 2011; Link et al, 2008). The DSF is a non-duplicative list of residential addresses where U.S. postal workers deliver mail; it includes city-style addresses and P.O. boxes, and covers single-unit, multi-unit, and other types of housing structures with known businesses excluded. In total the DSF is estimated to cover 97% of residences in the U.S., with coverage gradually increasing over the last few years as rural addresses are being converted to city-style, 911-compatible addresses.[1] The universe of sample units is defined as this set of residential addresses, and hence is capable of reaching all individuals who are 18 years of age or older living at a residential address in the 17 target states and the District of Columbia. Samples from DSF are taken indirectly, as USPS cannot sell mailing addresses or otherwise provide access to DSF. Instead, a number of sample vendors maintain their own copies of the DSF, and through verifying them with USPS, update

[1] For example, in rural areas, Rural Route box addresses have been converted to physical street addresses.

the list quarterly. The sample vendors can also augment the mailing addresses with additional information (household demographics, landline phone numbers, etc.) from external sources.

The sample was allocated in equal proportions of 33% for each stratum, thus leading to the highest sampling rate in the Bay States stratum, and the lowest sampling rate in the Other East Coast States stratum. Within each survey region, the household sample was allocated in proportion to the geographic distribution of households within states in the three regions. This allocation is designed to minimize the variance of the main effects under an experimental design model. As a result of stratification, the analysis produces estimates of the geographic distribution of values for the Chesapeake Bay water quality improvements with greater precision.

The sample size across all three geographic strata was 6,600 households; 2,200 addresses were randomly selected from the states in each stratum. This sample size was chosen to provide statistically robust regression modeling while minimizing the cost and burden of the survey. Given this sample size, the level of precision achieved by the analysis is more than adequate to meet the analytic needs of the benefits analysis for the Chesapeake Bay TMDLs.

The Dillman Tailored Design Method (TDM) was used to implement the survey as a mailed choice questionnaire (Dillman 2008). All mailings were addressed to "Current (State) Resident," where the state in which the resident resides was included in the address (e.g., Current Maryland Resident). This increased the likelihood that a mailing would be delivered to the household.[2] Five mailings were sent to each randomly selected resident, as indicated in Table 3. Each survey included a stamped, return envelope. And all mailings included phone and email contact information for questions.

Table 3 Survey Materials Mailing Schedule

Type of Mailing	Date
Preview Letter	May 9, 2014
Survey 1	May 16, 2014
Reminder Postcard	May 28, 2014
Survey 2	June 10, 2014
Reminder Letter	June 17, 2014

Prior to mailing the survey, households were sent a preview letter notifying them that they had been selected to participate in the survey and briefly describing the purpose of the study. The survey with a cover letter and self-addressed, stamped envelope was mailed one week later, followed by a reminder postcard. A second survey was mailed to those who did not respond to earlier mailings. A final reminder letter was mailed after the second survey. All letters told households whether or not they were located in the watershed and described the purpose of the survey.

[2] In the pre-test the name of the randomly selected resident was used on all mailings. If the resident had moved, the mailing was returned as undeliverable. However, if the mailing is addressed to "Current (State) Resident" it is delivered to the address. This increases the number of mailings that are delivered, but the lack of personalization could reduce responses if residents ignore the mailings.

EPA also conducted a non-response follow-up survey with a sample of households that did not respond to the main survey. A sample of 1,750 addresses was randomly selected from the set of households for which no contact was made in the main survey. Contact was made when a survey was returned as partially or fully completed, or due to a delivery issue (e.g., vacant household, no mail receptacle). The non-response sample was stratified using the same method as was employed in the main survey and was mailed on July 17, 2014.

Data quality was monitored by checking submitted surveys for completeness and consistency. Survey responses were entered into an electronic database after they are returned and EPA cleaned the data to ensure it was entered in a consistent manner and any discrepancies were addressed. The Double Entry data entry method for closed-ended responses was used. This method consists of data being keyed twice and compared. Discrepancies are reconciled upon completion of the second entry.

5. Data Summary

This section presents the results of the data collection efforts for the main survey and non-response follow up. While the constant baseline is treated as our base case and the improving and declining versions are analyzed separately, we combine the results of the data collection here when presenting response rates, demographic data, and responses to behavioral and attitudinal questions.

5.1 Response Rates

To calculate the response rate we use the American Association of Public Opinion Research's Response Rate 3 (RR3) which estimates what proportion of cases of unknown eligibility is actually eligible (Halpern et al. 2005). Specifically, the eligibility rate $e = \frac{I+R}{I+R+X}$, where *I* is the number of returned completed surveys, *R* is the number of refusals (phone calls, emails, or survey materials mailed back indicating unwillingness to participate) and *X* is the number of known ineligible cases (e.g. survey materials returned by the post office). The response rate is calculated as $RR3 = \frac{I}{I+R+eU}$, where *U* is the number of cases with unknown eligibility. Table 4 shows the values used to calculate RR3 and the result.

Table 4 Summary of Response Rates by Sampling Strata

	Bay States	Watershed States	Other Eastern States	Overall
Surveys Mailed	2,828	1,868	1,868	6,600
Undeliverable	174	92	150	414
Completed Surveys Returned	794	479	369	1,642
Estimated Eligible Rate	82.3%	84.3%	72.3%	80.4%
Response Rate (RR3)	34.1%	30.1%	27.1%	31.0%

A total of 6,600 addresses were included in the sample with more addresses selected in the Bay States than in the other two strata in order to administer the improving baseline version of the survey in that stratum only. The estimated eligibility rate ranges from 72.3% in Other East Coast States to 84.3% in the Watershed States and is 80.4% overall. As expected, because of proximity to the Chesapeake Bay, 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%.

5.2 Sample Characteristics

The survey included questions about the respondents and their households to allow comparisons with the sample frame population, evaluate the representativeness of our sample, and control for demographic variables when estimating WTP. Table 5 presents the characteristics of our main survey respondents, non-response follow up (NRFU) respondents, and the sample frame population.

Table 5 Demographic Comparisons

	Main Survey	Non-response Follow up	Sample Frame Population[3]
Male	52%	49%	48%
Hispanic	5%	7%	13%
Black	12%	15%	18%
Pacific Islander	0%	2%	<1%
Asian	4%	7%	5%
White	84%	77%	73%
4-Year College Degree	53%	44%	33%
Income			
Under $25,000	14%	16%	24%
$25,000-$49,999	17%	17%	24%
$50,000-$74,999	17%	13%	18%
$75,000-$99,999	14%	13%	12%
$100,000-$149,999	14%	12%	12%
$150,000-$199,999	7%	4%	5%
$200,000 or more	6%	7%	5%

The main survey sample contains fewer minorities than the sample frame population. It also contains a higher proportion of people with at least a four-year college degree and a higher income distribution. The NRFU sample exhibits the same trends but to a lesser extent in most respects. This suggests that the monetary incentive and shorter questionnaire of the NRFU was successful in recruiting people fitting demographic profiles that are less likely to complete this type of survey.

5.3 Awareness, Behavior, and Attitudes

The survey included several questions to probe the respondents' familiarity and experience with the Chesapeake Bay and freshwater lakes in the watershed. These questions were included in the non-response follow-up to compare samples and collect data that can be used to condition WTP estimates on variables for which we think the NRFU provides more representative responses. Table 6 summarizes responses to questions about peoples' familiarity and experiences with the Chesapeake Bay and Watershed lakes. As one might expect, respondents to the main survey are more likely to be familiar with the Chesapeake Bay and more likely to have visited the Bay or a freshwater lake in the watershed than NRFU respondents. Respondents to the main survey were also more likely to be aware of nutrient and sediment pollution than respondents to the NRFU.

[3] Source: US Census Bureau American Fact Finder

Table 6 Familiarity with Chesapeake Bay and Watershed Lakes

	Main Survey	Non-response Follow up
Heard of the Chesapeake Bay	94%	85%
Visited the Bay for recreation in the past 5 years	43%	28%
Visited a watershed lake for recreation in the past 5 years	39%	24%
Aware of nutrient and sediment pollution	80%	67%

Following the policy choice questions the survey asked a number of debriefing questions using a Likert scale response format. Each question asked the respondent to indicate how much they agreed or disagreed with a given statement on a scale of 1 to 5, from strongly agree to strongly disagree.   Table 7 summarizes the responses to several of those questions about the respondents' attitude toward different aspects of policies aimed at improving water quality. When asked about the statement that "It is important to improve waters in the Chesapeake Bay no matter how high the cost," 24.4% of the main survey respondents agreed (answering 1 or 2) compared with 11.4% of respondents to the non-response follow up. Nearly 40% of respondents to the main survey disagreed with the statement (answered 4 or 5) compared to just under half of the NRFU respondents. 47.2% of main survey respondents and 31.5% of respondents to the NRFU indicated they were against more regulations and government spending. Nearly equal proportions of the main survey and NRFU disagreed with the same statement. Finally, 36.4% of the main survey respondents and 18% of the NRFU respondents felt that that they should not have to pay to improve water quality in the Chesapeake Bay and Watershed lakes. 28.1% of main survey respondents and 42.8% of the NRFU respondents disagreed with that statement.

Table 7 Responses to Attitudinal Debriefing Questions

		Strongly Agree 1	2	3	4	Strongly Disagree 5	Don't Know
It is important to improve waters in the Chesapeake Bay no matter how high the cost	Main Survey	13.3%	11.1%	25.5%	19.5%	18.7%	4.9%
	NRFU	5.3%	6.1%	28.1%	23.6%	23.9%	11.0%
I am against any more regulations and government spending	Main Survey	23.6%	13.6%	21.4%	10.5%	16.8%	6.1%
	NRFU	17.1%	14.4%	28.0%	15.2%	11.7%	11.0%
My household should not have to pay to improve Bay waters and Watershed Lakes	Main Survey	20.9%	15.5%	23.1%	8.4%	19.7%	4.3%
	NRFU	9.5%	12.5%	23.7%	12.9%	29.9%	9.1%

6. Data Analysis

The survey data can be used to estimate the public's WTP for a range of water quality improvements. This section describes the econometric model we use to estimate the utility parameters that enter the WTP calculations, presents the regressions results, and presents the results of the non-response bias study.

6.1 Empirical Model

The empirical model is grounded in Random Utility Model (RUM) theory, which posits that utility is composed of a deterministic component, $U(\cdot)$, and an unobserved random component ε. In the base model 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 $v(\cdot)$ 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 (income minus the cost of alternative j, $(Y_i - C_j)$). The vector $\boldsymbol{x_j}$ includes attributes of the Bay (water clarity, bass, crab, and oyster populations) and the number of watershed lakes with low algae levels. Utility also depends on a stochastic component that is not observable to the researcher (ε_{ij}).

For choice occasion (or choice question) *t*, household *i* will choose alternative *j* if it yields the greatest utility over all other alternatives. More formally:

$$v_{itj} + \varepsilon_{itj} > v_{itk} + \varepsilon_{itk}, \forall\, k \neq j \qquad (2)$$

Assuming ε follows a type I extreme value (Gumbel) distribution, the model can be estimated as a conditional logit model, as detailed by Maddala (1983) and Greene (2003).

The literature offers no firm guidance regarding the choice of specific functional forms for *v(·)* within choice experiment estimation, but in practice linear forms are often used (Johnston et al. 2003).[4] Omitting the choice question subscript *t* for notational ease, and assuming a linear indirect utility function, the most basic model is:

$$u_{ij} = \boldsymbol{\beta} \boldsymbol{x}_j + \psi\left(Y_i - C_{tj}\right) + \varepsilon_{ij} \qquad (3)$$

where $\boldsymbol{\beta}$ denotes the marginal utility for a change in the environmental attributes and ψ is the marginal utility of income. When calculating the probability that respondent *i* chooses alternative *j*, the un-interacted income term Y_i drops out. Therefore, the most basic empirical model is:

$$u_{ij} = \boldsymbol{\beta} \boldsymbol{x}_j + \gamma C_j + \varepsilon_{ij} \qquad (4)$$

where γ is the negative of the marginal utility of income $(\gamma = -\psi)$. The parameters to be estimated are $\boldsymbol{\beta}$ and γ.

From the estimated parameters, marginal willingness to pay (MWTP) for a given environmental attribute can be calculated by taking the ratio of the marginal utility with respect to the environmental attribute and the marginal utility of income. More formally:

$$MWTP = \frac{\partial u_{ij} / \partial \boldsymbol{x}_j}{\partial u_{ij} / \partial C_j} = -\frac{\boldsymbol{\beta}}{\gamma} \qquad (5)$$

where the negative sign is needed since $\gamma = -\psi$. Equations (1) – (5) implicitly assume that the utility obtained from choosing an alternative, and the marginal utilities corresponding to the environmental attributes and income, are constant across households. Such assumptions are subsequently relaxed when estimating the empirical models by adding a series of interaction terms with household characteristics, as discussed in the next section.

6.2 Model Specification

Great care was taken in designing the survey instrument and in screening the sample in order to minimize potential biases (see sections 3.2 and 6.3, respectively). Nonetheless, it is still possible that some households exhibited responses reflecting protest, hypothetical bias, or "warm-glow" behaviors. Some respondents could be systematically choosing one alternative, such as the status quo, more or less often, irrespective of the environmental and cost attribute levels specified. Such residual behaviors, if present and not properly controlled for, would bias estimates of $\boldsymbol{\beta}$ and γ.

To test and control for such potential effects, an alternative specific constant is added to the status quo alternatives. If this constant term is positive and statistically significant, it would suggest respondents favor the status quo irrespective of the alternative improvements and costs; perhaps reflecting protest responses or "cold feet" towards a policy option. In contrast, a negative and significant

[4] Some apply more flexible forms, such as quadratic to allow for nonlinearities over the attribute space (Cummings et al. 1994). This was explored in earlier estimation procedures, but ultimately not pursued in the main analysis.

status quo constant (SQC) would suggest that respondents favor a policy option, but their choices are not necessarily linked to the actual environmental improvements described. 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. In order to allow for heterogeneity across respondents with regards to both possibilities, the base model and subsequent specifications include a SQC that is allowed to be stochastic following an assumed normal distribution. Mansfield et al. (2011) took a similar approach in their stated preference study of the Klamath River Basin.

Respondent heterogeneity is further controlled for in subsequent models by including interaction terms between household specific characteristics and the environmental and cost attributes.[5] Such characteristics include whether respondents are a user or non-user of the Chesapeake Bay and lakes in the watershed, whether they are located within or outside of the watershed, household annual income, and other sociodemographic characteristics.

6.3 Sample Weights and Screening

The objective of the analysis is to obtain an unbiased estimate of willingness to pay (WTP) that is representative of the population of households in the sample frame. In order to do so, prior to estimating the RUM regressions two procedures must take place: sample weighting and screening.

Sample Weights

As discussed in section 4, the survey was administered across three geographic strata: the "Bay States", other "Watershed States", and all other "East Coast States". In addition to different sampling intensities across the strata, there were differences in response rates, with a lower response rate in strata that are located farther from the Chesapeake Bay. In the regression analysis responses across all strata are pooled and each response is weighted to ensure households in each of the strata are sufficiently represented.

Observations are weighted following a two-step procedure. First, comparing the total population within each strata to the number of households sent the survey in each strata, the base weight is calculated as the inverse of the probability of a household being sent a survey. Second, strata weights are adjusted to account for differences in response rates across strata. Using the AAPOR3 response rates, which account for differences in ineligible addresses that could not receive a survey, the final sample weights are shown in table 9.

Table 8 Base Sample Weights across Geographic Strata

Stratum	Total Households	Households sent Constant Baseline Surveys	Base Weight
Bay States	5,479,176	943	5,816.535

[5] 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).

Watershed States	13,442,787	943	14,270.475
Other East Coast States	25,431,478	943	26,997.323

Table 9 Final Sample Weight after Adjusting for Response Rate

Stratum	Base Weight	Response Rate	Response Weight
Bay States	5,816.535	0.341	17,057.2874
Watershed States	14,270.475	0.301	47,410.2159
Other East Coast States	26,997.323	0.271	99,621.1181

Screening the Sample

In order to avoid biases associated with protest responses, scenario rejection, and hypothetical bias responses to both the choice scenarios and debriefing questions were screened to identify respondents who exhibited such behaviors. Such behaviors are likely to bias results (either in a positive or negative direction, depending on the behavior) and therefore are excluded from the analysis.

Respondents identified as exhibiting protest responses of different degrees were identified as:

Protest 1

- Respondent always chose status quo in the choice questions, and
- Agreed or strongly agreed to the statement that “I am against any more regulations and government spending”, and
- Agreed or strongly agreed to the statement that “My household should not have to pay any amount to improve Bay Waters and Watershed lakes.”

Protest 2

- Respondent always chose status quo in the choice questions, and
- Agreed or strongly agreed with at least one of the following statements:
    - “I am against any more regulations and government spending”, or
    - “My household should not have to pay any amount to improve Bay Waters and Watershed lakes”

Other behaviors are expected to bias WTP upward. Respondents identified as exhibiting warm-glow and hypothetical bias are those who:

Warm-glow

- Always chose most expensive option in the choice questions, and
- Agreed or strongly agreed to the statement that “It is important to improve waters in the Chesapeake Bay Watershed, no matter how high the costs.”

Hypothetical bias 1

- Respondents always chose most expensive option in the choice questions, and
- Disagreed or strongly disagreed with the statement that "I voted as if my household would actually face the costs shown in the questions."

Hypothetical bias 2

- Respondents who were identified under *Hypothetical 1*, or
- Respondents disagreed or strongly disagreed with the statement that "I voted as if the programs would actually achieve the results shown by 2025"

These different types of biases are treated symmetrically in the analysis by screening out respondents who exhibited such behaviors, and then estimating the random utility models based on the remaining sample[6]. Table 10 displays the number of respondents remaining under alternative screening criteria. These criteria get more stringent moving left to right, and top to bottom. The original unscreened sample includes 684 respondents. The main regression results and WTP estimates in this report are based on the 523 respondents who do not provide evidence of protest, warm glow, and hypothetical bias, as identified above responses (bottom right cell). Sensitivity of our results to these screening criteria are included in section 8.

Table 10 Number of respondents screened by alternative criteria (Constant Baseline Survey)

	Unscreened	Warm-glow & Hypothetical Bias 1	Warm-glow & Hypothetical Bias 2
Unscreened	684	678	630
Protest 1	620	614	578
Protest 2	559	553	523

6.4 Regression Results

The base RUM regression coefficient results (model 1) are presented in column 1 of table 11. As one might expect 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 statistically significant coefficient on the cost attribute implies a positive marginal utility of income.

The SQC is allowed to randomly vary across respondents, but the negative and significant mean value (*status quo*) shows that, on average, respondents are more likely to choose a policy option (option

[6] Banzahf et al. (2006) take a similar approach to show the effect of screening out responses flagged by yea-saying and nay-saying indicators on estimated WTP. Thanks to Alan Krupnick for this suggestion.

b or option c) over the status quo (option a). The statistically significant estimated standard deviation corresponding to the SQC shows that there is unobserved variation across respondents regarding the SQC, validating the base model specification.

Table 11 RUM Regression Coefficient Results. All Strata, Constant Baseline Survey

VARIABLES	Model 1 (Base Model)	Model 2	Model 3
clarity	2.266e-01*	5.687e-02	5.957e-02
	(0.126)	(0.154)	(0.161)
bass	4.052e-02**	4.271e-02**	4.278e-02**
	(0.017)	(0.020)	(0.021)
crab	8.755e-03***	8.910e-03***	9.391e-03***
	(0.002)	(0.003)	(0.003)
oysters	5.704e-05**	3.405e-05	4.424e-05
	(0.000)	(0.000)	(0.000)
lakes	1.210e-03***	1.130e-03***	1.153e-03***
	(0.000)	(0.000)	(0.000)
clarity × user		5.570e-01**	5.748e-01**
		(0.244)	(0.262)
bass × user		-1.699e-03	2.618e-04
		(0.031)	(0.033)
crab × user		-1.066e-03	5.455e-04
		(0.004)	(0.004)
oysters × user		8.692e-05	1.145e-04*
		(0.000)	(0.000)
lakes × user		3.760e-04	4.320e-04
		(0.000)	(0.000)
clarity × watershed			-2.568e-02
			(0.245)
bass × watershed			-4.292e-03
			(0.030)
crab × watershed			-5.298e-03
			(0.004)
oysters × watershed			-1.017e-04*
			(0.000)
lakes × watershed			-2.199e-04
			(0.000)
cost	-7.400e-03***	-7.532e-03***	-7.555e-03***
	(0.001)	(0.001)	(0.001)
status quo (mean)	-1.749e+00***	-1.753e+00***	-1.753e+00***
	(0.358)	(0.357)	(0.353)
status quo (std dev)	4.127e+00***	4.048e+00***	4.028e+00***
	(0.473)	(0.463)	(0.463)

In Model 2, interaction terms are added between the Bay attributes (clarity, bass, crab, and oysters) and a dummy variable denoting active users of the Bay, as well as an interaction term between the watershed lakes attribute and a dummy variable identifying active users of watershed lakes. The coefficient estimates on these interaction terms are mixed. The only statistically significant estimate corresponds to *clarity × user*, which indicates that users of the Bay do hold a premium for clearer waters compared to non-users. The positive coefficients on the un-interacted environmental attributes, at least when statistically significant, suggest that non-users hold a positive marginal utility for improvements, particularly for bass, crab, and watershed lakes.

Additional interaction terms between the environmental attributes and a dummy variable denoting whether the household is located within the watershed are added in Model 3. 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 may be more intense users. The coefficients on these interaction terms are not statistically different from zero, neither individually nor jointly.[7] It seems that conditional on user status, households inside the watershed have similar preferences towards the Bay and lakes attributes as those outside the watershed. Notice that after controlling for location in or out of the watershed, the user interaction terms are all positive, providing some evidence that users do hold an additional premium for improvements in these attributes. The coefficients on the user interaction terms are not always significant, both individually and jointly (p=0.1238).

Following equation (5), the MWTP estimates from Models 1 through 3 are calculated and presented in table 12. All WTP estimates are reported in 2014 dollars. Looking first at model 1, we see that the MWTP estimates are positive and significant for all attributes. Estimated WTP for a one foot increase in bay water clarity is $30.62. Respondents are willing to pay $5.47 for to increase the population of striped bass by one million, $1.18 for a million additional blue crabs, and $0.01 for a one ton increase in oysters. Finally, an additional low algae lake is valued at $0.16 per respondent.

[7] The only exception is the negative and marginally significant coefficient on *oysters × watershed*. Nonetheless, a chi square test fails to reject the null that the coefficients are jointly insignificant (chi2(5)=5.95, p=0.3111).

Table 12 Marginal Willingness to Pay Estimates. All Strata, Constant Baseline Survey (2014 dollars).

	Model 1	**Model 2**		**Model 3**			
				In Watershed		Out of Watershed	
		Users	Nonusers	Users	Nonusers	Users	Nonusers
Clarity (feet)	30.62*	81.50***	7.55	80.57***	4.49	83.97***	7.89
	(16.90)	(26.50)	(20.41)	(27.14)	(32.54)	(32.33)	(21.20)
Bass (million fish)	5.47**	5.44	5.67**	5.13	5.09	5.70	5.66**
	(2.25)	(3.41)	(2.64)	(3.52)	(3.86)	(4.06)	(2.76)
Crab (million crab)	1.18***	1.04**	1.18***	0.61	0.54	1.32**	1.24***
	(0.32)	(0.44)	(0.39)	(0.50)	(0.57)	(0.54)	(0.40)
Oysters (tons)	0.01**	0.02***	0.00	0.01	-0.01	0.02***	0.01
	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
Lakes (# of lakes)	0.16***	0.20***	0.15***	0.18***	0.12***	0.21***	0.15***
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)	(0.03)

Model 2 allows for the comparison of MWTP between users and non-users. Users of the Bay hold a significantly higher MWTP for an improvement in clarity compared to non-users (chi2(1)=4.98, p=0.0257). Otherwise the MWTP are fairly similar, with the point estimates of users MWTP for bass and crab being slightly lower, and that for oysters and lakes being slightly higher. In all these cases, however, these differences are not statistically significant.

In Model 3 the MWTP of households within or outside the watershed can be separated and compared. The results suggest that MWTP for the different attributes is similar for respondents that are located within or outside of the Chesapeake Bay Watershed. Chi-square tests comparing the user or non-user MWTP between respondents in and out of the watershed reveal no statistically significant differences, although households outside the watershed hold a slightly higher MWTP for oysters (p=0.0596).

The next set of models are presented in table 13. Building off of Model 3, household heterogeneity is further accounted for by interacting the cost and/or environmental attributes with household income, education, race, and ethnicity. In Model 4 the marginal utility of income is allowed to vary across income groups. Respondents were divided into three annual income groups ($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, consistent with a decreasing marginal utility of income. However, in this specification the marginal utility of income is not statistically different across the three income groups (chi2(2)=3.85, p0.1462).[8]

Table 13 RUM Regressions with Additional Household Characteristics (Constant Baseline Survey)

VARIABLES	Model 4	Model 5	Model 6
clarity	1.074e-01	-1.824e-01	1.135e-01
	(0.168)	(0.214)	(0.228)
bass	4.634e-02**	4.276e-02	1.040e-01***
	(0.022)	(0.028)	(0.032)
crab	1.010e-02***	5.237e-03	8.623e-03*
	(0.003)	(0.004)	(0.005)
oysters	6.631e-05**	6.508e-05	1.379e-04***
	(0.000)	(0.000)	(0.000)
lakes	1.220e-03***	5.740e-04*	9.554e-04***
	(0.000)	(0.000)	(0.000)
clarity × user	5.313e-01*	4.119e-01	2.555e-01
	(0.277)	(0.283)	(0.295)
bass × user	3.729e-04	2.869e-03	-4.806e-03
	(0.036)	(0.037)	(0.040)
crab × user	1.389e-03	-6.292e-04	-1.094e-03
	(0.005)	(0.005)	(0.005)
oysters × user	1.064e-04*	1.058e-04*	1.128e-04
	(0.000)	(0.000)	(0.000)
lakes × user	3.597e-04	3.279e-04	3.718e-04
	(0.000)	(0.000)	(0.000)
clarity × watershed	-4.013e-02	-2.605e-02	2.340e-02
	(0.258)	(0.262)	(0.276)
bass × watershed	-7.258e-03	-8.865e-03	-2.122e-02
	(0.033)	(0.033)	(0.036)
crab × watershed	-6.466e-03	-5.471e-03	-6.987e-03

[8] Pairwise comparisons reveal that households in the $50,000 to $100,000 annual income category hold a slightly higher marginal utility of income compared to households with an annual income less than $50,000 or greater than $100,000 (p=0.0897 and p=0.0925, respectively). The marginal utility of income is not statistically different across the lowest and highest income categories (p=0.8918), however.   This more flexible categorical specification is preferred, but several alternative models were explored, including interactions between *cost* and a linear 1 to 7 scalar denoting the different income groups elicited in the survey, and a dummy variable denoting households with income levels above the median. These models constrain the relationship between marginal utility of income and household income to be monotonic; and in doing so, although not statistically significant, the positive sign of the coefficients on these interaction terms are consistent with the notion of a decreasing marginal utility of income.

VARIABLES	Model 4	Model 5	Model 6
	(0.005)	(0.005)	(0.005)
oysters × watershed	-1.137e-04**	-1.134e-04**	-1.129e-04*
	(0.000)	(0.000)	(0.000)
lakes × watershed	-2.340e-04	-2.416e-04	-2.585e-04
	(0.000)	(0.000)	(0.000)
clarity × degree		6.195e-01**	3.751e-01
		(0.253)	(0.269)
bass × degree		1.265e-02	-2.611e-02
		(0.033)	(0.036)
crab × degree		1.007e-02**	8.824e-03
		(0.005)	(0.006)
oysters × degree		1.142e-05	-5.232e-05
		(0.000)	(0.000)
lakes × degree		1.192e-03***	8.274e-04**
		(0.000)	(0.000)
cost × degree		-2.329e-04	2.158e-03
		(0.001)	(0.002)
clarity × black			-6.725e-01
			(0.639)
bass × black			-1.852e-01***
			(0.045)
crab × black			-1.348e-02
			(0.009)
oysters × black			-2.505e-04***
			(0.000)
lakes × black			-1.111e-03
			(0.001)
cost × black			8.781e-03***
			(0.002)
clarity × hispanic			-1.796e+00*
			(0.993)
bass × hispanic			-4.465e-02
			(0.077)
crab × hispanic			-2.610e-02***
			(0.009)
oysters × hispanic			-2.764e-04*
			(0.000)
lakes × hispanic			6.432e-04
			(0.001)
cost × hispanic			7.253e-03***
			(0.002)

VARIABLES	Model 4	Model 5	Model 6
cost × income 0-50k	-6.358e-03***	-5.974e-03***	-9.269e-03***
	(0.001)	(0.001)	(0.002)
cost × income 50-100k	-9.301e-03***	-9.271e-03***	-1.275e-02***
	(0.001)	(0.001)	(0.002)
cost × income 100k+	-6.589e-03***	-7.090e-03***	-1.025e-02***
	(0.001)	(0.002)	(0.002)
status quo (mean)	-1.795e+00***	-1.711e+00***	-1.870e+00***
	(0.380)	(0.359)	(0.405)
status quo (std dev)	4.128e+00***	3.806e+00***	3.907e+00***
	(0.525)	(0.434)	(0.469)
Observations	3,954	3,945	3,603
ll	-5.7368e+07	-5.6477e+07	-4.9725e+07

Model 5 adds interactions between cost and the environmental attributes with a dummy variable denoting respondents who have a Bachelor's degree or higher (*degree*). It is possible that such households have systematically different preferences, and it is important to account for such heterogeneity if it exists. The coefficients corresponding to the interaction terms between *degree* and the environmental attributes are all positive, and are statistically significant for *clarity*, *crab*, and *lakes*. These coefficients are also jointly significant (chi2(6)=19.40, p=0.0035). Together this suggests that education is an important factor in accounting for MWTP for improvements in the Chesapeake Bay and Watershed lakes.

Model 6 further allows for household heterogeneity by accounting for race and ethnicity. Interaction terms between all attributes and dummy variables denoting whether a respondent is Black or African American (*black*) and of Hispanic, Latino, or Spanish origin (*hispanic*) are added. The coefficients on these interaction terms are jointly, and often individually significant. Again suggesting that race and ethnicity are important factors in estimating a household's MWTP.[9, 10]

[9] Chi-square tests of the joint significance of the interaction terms between the attributes and *black* and *hispanic* yield: chi2(6)=49.47 (p=0.0000) and chi2(6)=67.28 (p=0.0000), respectively

[10] Additional models not reported here explored whether other aspects of household heterogeneity were important in modelling choices for improvements in the Bay and Watershed lakes. 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, both individually and jointly.

6.5 Non-response Analysis

Non-response bias occurs when a population characteristic is being estimated with survey data which, due to non-response, underrepresents certain types of respondents. In this case the population characteristic of interest is WTP for water quality improvements. While WTP may be correlated with demographic characteristics for which census data are available, such as education and income, it is also likely to be correlated with behavioral and attitudinal data that are not available elsewhere. The purpose of the non-response follow up (NRFU) questionnaire is to collect demographic, behavioral, and attitudinal data from households that did not return the main survey to evaluate the potential for non-response bias and collect data that can be used to condition the WTP estimates in order to better represent the general population.

Section 3.5 describes the questions included on the NRFU. Summary statistics and comparisons between the main survey sample, NRFU sample, and population of the sample frame are shown in the Data Summary section in tables 5, 6, and 7. In this section we analyze responses to behavioral and attitudinal questions that were included on the main survey and the NRFU to test for differences between the two samples.

Table 14 shows questions from the surveys that asked about respondents' familiarity and experience with the Chesapeake Bay and freshwater lakes in the watershed. The columns of table 14 show the percentage of respondents to each survey that answered in the affirmative and the t-statistic for a difference of means test. Respondents who completed and returned the main survey are more likely to be familiar with and to have visited the Chesapeake Bay than the NRFU sample, though the difference is not always statistically significant. This trend extends to experiences with freshwater lakes in the watershed as well.

Table 14 Comparing Familiarity and Experience Across Main Survey and NRFU Samples

	Main Survey	NRFU	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*

Table 15 summarizes responses to the attitudinal questions that followed the policy choice questions. These questions used a Likert scale response format with values from 1 = Strongly Disagree to 5 = Strongly Agree. The values shown in table 15 are the means of the numerical responses to each question. Respondents to the NRFU are more like to say that it is important to improve water quality regardless of costs but also that they should not have to bear the costs. The mean response to regulations and government spending is only slightly on the disagree side of the scale in both samples. Finally, NRFU respondents are more likely to say it is difficult to find time to take surveys.

Table 15 Comparing Attitudinal Responses across Main Survey and NRFU Samples

	Main Survey	NRFU	t-test of Means
It is important to improve waters in the Chesapeake Bay Watershed, no matter how high the costs	3.20	3.63	4.68***
I am against anymore regulations and government spending	2.84	2.89	0.45
My household should not have to pay any amount to improve Bay Waters and Watershed Lakes	2.89	3.47	5.36***
It is difficult for me to find time to take surveys	2.71	3.26	5.05***

Data from the main survey and the NRFU 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 NRFU sample. While only a few of these differences are statistically significant, their experiences could influence their responses to the policy choice questions and our WTP estimates. Data from the attitudinal questions, however, are even less conclusive. The two samples have very similar attitudes toward more regulation and government spending. NRFU 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.

In the data summary section the demographic characteristics of both samples were compared to the population of the sample frame. By every measure the NRFU sample was closer to the population than the main survey sample. For this reason we choose to use the NRFU proportions when conditioning WTP estimates on variables that are not available elsewhere; such as a proportion for users of the Bay and Watershed lakes.

7. Estimating Household and Total WTP

The preference parameter estimates from the RUM regressions are used to estimate the benefits of the environmental improvements projected to occur under the full implementation of the TMDL, incremental to the projected baseline scenario.[11] Details of the projected levels of different features of

[11] It is important to note that the projected baseline scenario in the broader benefit-cost analysis is different from the status quo option outlined in the survey instrument. This was done to increase future applicability of the survey results to other actions to improve the Chesapeake Bay and Watershed lakes, as well as to accommodate alternative ecological/hydrological models and assumptions. Since the attribute space covered in the survey

the environment under the baseline and TMDLs scenario are discussed in chapter *#TBD*. The projected changes for the attributes included in the choice questions are presented in table 16.

Table 16 Projected Changes in Environmental Attributes

Environmental Attribute	Projected Change Under TMDL Relative to Projected Baseline
Bay Water Clarity	+0.361 feet
Striped Bass Populations	+0.312 to 1.032 million fish
Blue Crab Populations	+2 to 41 million crab
Oyster Populations	+6.6 to 541.2 tons
Low Algae Watershed Lakes	+455 lakes

The ranges for the projected changes in striped bass, blue crab, and oyster populations arise from alternative projections, where the lower end is based on a multi-species model and the higher end is based on expert elicitation (See chapter *#TBD* for details on the ecological projections). The monetary value for these projected improvements are calculated in the following sections.

7.1 Estimating Household WTP

Welfare calculations stemming from RUM results are relatively well developed in the literature (Hanemann, 1999; Morey, 1999). Following the approach outlined by Holmes and Adamowicz (2003, pages 194-197), household WTP is calculated as:

$$WTP^{HH} = \frac{(v^1 - v^0)}{-\gamma} \qquad (6)$$

where v^1 and v^0 are the deterministic components of utility evaluated at the projected attribute levels under the TMDL and baseline scenarios, respectively.[12] Further, if $v(\cdot)$ is assumed to be linear, as is the case in equations (3) and (4), the household WTP expression in equation (6) can be re-written as the MWTP for each attribute multiplied by the projected change in that attribute.

As an illustration, table 17 displays the household WTP estimates based on model 3 (from table 11), 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, and also distinguishes between households located in or outside of the watershed. The WTP estimates for all household types are

design encompasses the range of attribute levels projected to occur under the baseline and TMDL scenarios in the broader benefit-cost analysis, applying the RUM results from the conjoint choice experiment in this fashion is valid.

[12] 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 v^1 and v^0. This is equivalent to assuming zero income effects (Holmes and Adamowicz, 2003).

positive and significant. This holds whether the WTP estimates are based on the high or low fish/shellfish projections. The table also shows that users of the Bay and Watershed lakes tend to have a higher WTP.[13] Estimates of WTP are nominally higher for households outside of the watershed but nonlinear Wald tests fail to reject the null that the annual corresponding user and non-user WTP estimates are equal (p-values ranging from p=0.1514 to p=0.5226).

Table 17 Annual Household WTP for TMDL (from Model 3 in 2014 dollars)

Fish/shellfish	In Watershed		Out of Watershed	
Projections	Users	Nonusers	Users	Nonusers
High	145.859*** (31.191)	81.174** (35.639)	196.955*** (37.625)	132.270*** (27.942)
Low	114.190*** (23.064)	60.449** (25.093)	130.328*** (26.285)	76.587*** (18.489)

Distinguishing between different household types can get fairly complicated, particularly in later models where other dimensions of heterogeneity are accounted for, including: household income, education, race, and ethnicity. Therefore, when calculating total benefits an alternative approach is taken, where WTP is calculated for a "representative" household. Similar to the above exercise, the coefficient estimates from the RUM results are used, but instead of plugging in household specific values for each variable, the population proportions are entered instead. This provides an average household WTP, weighted to reflect the distribution of users, income, education, etc. within the population of the study frame. A similar approach is taken by Van Houtven et al. (2014) in their analysis of lake water quality in Virginia.

To illustrate this representative household WTP calculation, again consider Model 3. Following equation (6) and the functional form assumed in equation (4), the representative household WTP (WTP^{RHH}) for model 3 is calculated as:

$$WTP^{RHH} = \frac{\{\hat{\boldsymbol{\beta}}+(\hat{\boldsymbol{\beta}}_{\boldsymbol{user}}\times\%\,\boldsymbol{user})+(\hat{\boldsymbol{\beta}}_{\boldsymbol{wshed}}\times\%\,\boldsymbol{wshed})\}\boldsymbol{x}^{\mathbf{1}}-\{\hat{\boldsymbol{\beta}}+(\hat{\boldsymbol{\beta}}_{\boldsymbol{user}}\times\%\,\boldsymbol{user})+(\hat{\boldsymbol{\beta}}_{\boldsymbol{wshed}}\times\%\,\boldsymbol{wshed})\}\boldsymbol{x}^{\mathbf{0}}}{-\hat{\gamma}}$$

$$= \frac{\{\hat{\boldsymbol{\beta}}+(\hat{\boldsymbol{\beta}}_{\boldsymbol{user}}\times\%\,\boldsymbol{user})+(\hat{\boldsymbol{\beta}}_{\boldsymbol{wshed}}\times\%\,\boldsymbol{wshed})\}\times(\boldsymbol{x}^{\mathbf{1}}-\boldsymbol{x}^{\mathbf{0}})}{-\hat{\gamma}} \quad (7)$$

where $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}_{\boldsymbol{user}}$, and $\hat{\boldsymbol{\beta}}_{\boldsymbol{wshed}}$ are the estimated coefficient vectors corresponding to the environmental attributes and the interaction terms between these attributes and the dummy variables denoting users and households who live within the watershed, respectively. The estimated marginal utility of income is

[13] The household WTP estimates for users and nonusers are statistically different from each other at conventional levels, with p-values ranging from p=0.0426 to p=0.0844.

denoted by $-\hat{\gamma}$, and $\boldsymbol{x^1}$ and $\boldsymbol{x^0}$ are the projected attribute levels under the TMDL and baseline scenario, respectively.

The proportions of the population that are users of the Bay and Watershed lakes (*% user)* and who live in the watershed (*% wshed*) are plugged into equation (7), according to the values presented in table 18. Similar calculations are conducted for more complex models that also account for income, education, race, and ethnicity, using the population values displayed below.

Table 18 Population Statistics Used in Estimating WTP

Variable	% Population	Source
Bay User	15.64%	Nonresponse follow-up survey
Lake User	14.62%	
In Watershed	15.12%	2010 US Census.
Annual Income $0 - $50k	49.365%	US Census 2009 American Communities Survey
Annual Income $50 - $100k	30.290%	
Annual Income >$100k	20.345%	
College Degree or Higher	48.1%	2010 US Census
Black or African American	18.2%	
Hispanic, Latino, or Spanish Origin	4.2%	

The annual WTP for the representative household is calculated using each of the main models presented in section 6.4. These models are progressively more flexible, allowing for increasing dimensions of household heterogeneity by introducing various interaction terms between the household characteristics and the environmental and cost attributes. The representative household WTP estimates are presented in table 19, under both sets of projected changes in fish and shellfish populations. The annual WTP estimates for the representative household are fairly robust across the different models, ranging from $133 to $167 under the high projected improvements in fish/shellfish populations, and $82 to $107 under the low projections.

Table 19 Annual WTP Estimates for the Representative Household (2014 dollars)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
High fish/shellfish projections	143.790*** (23.632)	135.302*** (24.458)	134.396*** (24.207)	166.707*** (35.467)	144.104*** (28.253)	133.045*** (28.265)
Low fish/shellfish projections	89.591*** (15.806)	82.592*** (16.170)	82.286*** (16.019)	107.513*** (27.662)	85.421*** (18.099)	82.563*** (20.031)

7.2 Estimating Total WTP

The representative household WTP values are then multiplied by the corresponding number of households in the population of our study frame. 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 frame. A critical consideration, however, is what to assume 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 care much about improvements to the Chesapeake Bay. 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. In the former case, the representative household WTP estimates would only be applied to the proportion of the population corresponding to the proportion of respondents (i.e., response rate × total number of households). In the latter case, the household WTP estimates would be 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 (Van Houtven et al., 2014; Mansfield et al. 2011).

As described by the nonresponse bias analysis in section 6.5, on average respondents to the main survey (compared to those who responded to the NRFU and presumably to the population) are more likely to have heard of the Chesapeake Bay, visited it, and were more aware of nutrient and sediment pollution issues. There is also some evidence that the survey sample had a disproportionately higher number of households who frequently see and participate in recreational activities in watershed lakes. While only a few of these differences are statistically significant, households' experiences could be positively correlated with their WTP towards improvements in the Chesapeake Bay and Watershed lakes. Although the WTP estimates are conditional on some of these differences, and are adjusted to reflect proportions in the population (e.g., user versus non-user), it is possible that some key differences are not controlled for.

Overall, there is no clear conclusion that can be drawn about how similar the preferences are between respondents and non-respondents. Hence WTP for Bay and Watershed lake improvements for non-respondents may be similar to respondents or may be lower, perhaps even close to zero. Given these uncertainties, two extreme cases are adopted in order to provide an upper and lower bound on the total annual WTP. As an upper bound, it is assumed that the sample is fully representative of the population of households, and total annual WTP is calculated as:

$$Total\ WTP^{Low} = WTP^{RHH} \times 44{,}353{,}441\ households \qquad (8)$$

As a lower bound, the WTP for the proportion of households that did not respond to the main survey is disregarded, and so the total annual WTP is calculated as:

$$Total\ WTP^{High} = WTP^{RHH} \times 44{,}353{,}441\ households \times 31.0\% \qquad (8)$$

The 31.0% corresponds to the overall response rate to the main survey. The upper and lower bound total annual WTP estimates are presented for each model in tables 20 and 21.

Table 20 Total Annual WTP Estimates (billions of 2014 dollars): WTP of population assumed to be similar to respondents

Fish/shellfish projections	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
High	6.378*** (1.048)	6.001*** (1.085)	5.961*** (1.074)	7.394*** (1.573)	6.392*** (1.253)	5.901*** (1.254)
Low	3.974*** (7.011)	3.663*** (7.172)	3.650*** (7.105)	4.769*** (1.227)	3.789*** (8.028)	3.662*** (8.885)

Table 20 Lower Bound Total Annual WTP (billions of 2014 dollars): WTP of non-respondents assumed to be zero

Fish/shellfish projections	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
High	1.977*** (3.249)	1.860*** (3.363)	1.848*** (3.328)	2.292*** (4.877)	1.981*** (3.885)	1.829*** (3.886)
Low	1.232*** (2.173)	1.136*** (2.223)	1.131*** (2.203)	1.478*** (3.803)	1.174*** (2.489)	1.135*** (2.754)

Conditional on the upper or lower bound assumptions for the proportion of non-responding households with nonzero WTP and on the fish/shellfish projections, WTP estimates are fairly robust across model specifications. No clear trend in total WTP is apparent with increased model complexity. Model 6 provides estimates at the lower end of the range, but this most complex model best reflects the distribution of household characteristics in the study frame population, and in that sense provides the most accurate estimate of total annual WTP, ranging from $1.135 billion to $5.901 billion (with a 95% confidence interval spanning $0.60 billion to $8.36 billion).

8. Validity Tests and Sensitivity Analysis

The first portion of this section contains several tests to examine the validity of the survey instrument and the resulting empirical estimates. These validity tests include several variants of tests for scope, assessing the consistency of preferences across choice questions, and an examination of relationships predicted by economic theory. The remainder of the section then provides a sensitivity analysis of household and total willingness to pay (WTP) estimates by varying different analytical assumptions, including: the identification and screening of respondents exhibiting protest, warm-glow, and hypothetical biases, consideration of omitted variables and baseline conditions.

Given that WTP estimates are fairly stable across Models 1-6, the validity tests and sensitivity analyses will be performed on the base model (Model 1) for ease of presentation and interpretation.

8.1 Validity Tests

Internal and External Scope

A common validity test of SP studies is examining whether the results pass the "scope" test. Scope tests assess whether stated responses show that utility is increasing with the quality or quantity of the amenity of interest. If so, this provides support that the respondents' are treating the hypothetical choice questions in ways similar to an actual market transaction and are making choices consistent with economic theory. As can be seen in our base model in column 1 of table 22, the coefficients on all of the environmental attribute are positive and statistically significant. Therefore, in this most basic sense the results pass the scope test.

A more stringent test is that for external scope. Even if the results pass the above scope test, this does not necessarily imply that choices stem from stable underlying preferences for the environmental commodity. It could be that respondents are merely being internally consistent, in that they anchor to, or rationalize, their first response by answering in a consistent manner to subsequent questions (see Ariely et al., 2003). A more robust test is the external scope test, where the researcher examines whether the coefficients are significant and of the expected sign when estimated from a cross-section of responses, where only one choice question per respondent is included. In this case, internal consistency or rationalization within a respondent is not possible, since only variation across respondents is used to estimate the coefficients. If the coefficient estimates are statistically significant and of the expected sign, then the results pass the external scope test (Carson 1997), and this provides further evidence that respondents are not just being internally consistent, but that their stated choices are consistent with economic theory, and thereby lend support to the validity of the survey instrument.

Columns 2 through 4 of table 22 present the results of RUM regressions that are estimated using only the first, second, and third choice questions, respectively. The estimated coefficients in these models only reflect variation across respondents. With the exception of the bass coefficient in the model examining the third question, the coefficient estimates are all of the expected sign and are often statistically significant, thus providing support that the results pass the external scope test.

Table 21 RUM Regressions Testing for Internal and External Scope

VARIABLES	Model 1	First Question Only	Second Question Only	Third Question Only
clarity	2.266e-01*	3.706e-01	2.767e-01	3.503e-02
	(0.126)	(0.252)	(0.255)	(0.187)
bass	4.052e-02**	7.846e-02**	5.589e-02**	-1.738e-02
	(0.017)	(0.036)	(0.028)	(0.021)
crab	8.755e-03***	1.177e-02**	1.203e-03	2.503e-03
	(0.002)	(0.005)	(0.006)	(0.003)
oysters	5.704e-05**	1.674e-04**	1.158e-05	1.811e-05
	(0.000)	(0.000)	(0.000)	(0.000)
lakes	1.210e-03***	9.762e-04**	1.185e-03***	8.651e-04***
	(0.000)	(0.000)	(0.000)	(0.000)
cost	-7.400e-03***	-1.084e-02***	-6.617e-03***	-4.747e-03***
	(0.001)	(0.002)	(0.001)	(0.001)
status quo (mean)	-1.749e+00***	-2.942e+00**	-5.565e-01	-7.195e-01**
	(0.358)	(1.386)	(0.356)	(0.335)
status quo (std dev)	4.127e+00***	5.828e+00***	-1.185e+00	4.196e-01
	(0.473)	(2.245)	(1.229)	(0.969)
Observations	4,311	1,467	1,464	1,380
ll	-6.2776e+07	-2.4621e+07	-2.3624e+07	-2.3208e+07

Although not a test in the formal sense, one can also examine scope sensitivity. In other words, are the magnitude of the coefficients and resulting MWTP estimates reasonable? There is no theoretical guidance on the degree of scope sensitivity needed in order for a study to be considered valid. Comparable revealed preference data could inform expectations, but such data are usually unavailable and therefore no conclusive “test” can be performed. One can, however, informally assess the magnitude of the estimates and make a judgement as to whether the differences in magnitude across different changes in the attributes are reasonable. For example, if respondents were willing to pay $10 a year to improve 5 lakes but only $15 per year to improve 500 lakes, this may lead to some skepticism in the results (see Bateman and Brouwer 2005).

To examine scope sensitivity the base model was re-estimated where the continuous attribute measures were replaced with dummy variables denoting each attribute level. The lowest (status quo) level for each attribute is omitted for statistical identification. Dividing the resulting coefficients by the marginal utility of income yields the WTP to go from the status quo level to some higher level for that attribute. The estimates are displayed in table 23.

The results show that respondents are willing to pay more to obtain a higher level for each of the attributes, and the differences in these WTP figures seem plausible. For example, respondents have a willingness to pay to improve water clarity by ½ a foot (from 3 to 3.5 feet) that is not statistically different from zero, whereas to the WTP to improve water clarity by 1.5 feet (3 to 4.5) is just under $45 per year. Consider the number of low algae watershed lakes as another example, respondents are willing to pay $120 to have 400 lakes switch from high to low algae levels (2,900 to 3,300 lakes), whereas the WTP to obtain a switch of 550 lakes (3,300 to 3,850) is $152. This additional $32 per year (=$152-$120) for an additional 550 low algae lakes seems plausible, considering the $120 WTP for the first 400 lakes. Similar trends hold for the striped bass and crab attributes. The only exception is the oysters attribute, where the point estimates suggest that respondents are willing to pay more for a smaller increase in oyster populations. Perhaps presenting oyster populations in units of tons was too abstract and more challenging for respondents compared to actual population figures, which were the metric for bass and crabs. Although, it is worth noting that the oyster WTP estimates are not statistically different from each other.

Table 23 RUM Regressions Testing for Scope Sensitivity

	Change	WTP (2014 Dollars)	H0: WTPs Equal
clarity	3 to 3.5 feet	-10.07 (20.99)	
	3 to 4.5 feet	44.79* (24.12)	chi2(1)=5.80** p=0.0160
bass	24 to 30 mil fish	21.22 (23.87)	
	24 to 36 mil fish	82.82*** (27.34)	chi2(1)=4.89** p=0.0270
crab	250 to 285 mil crab	24.46 (21.02)	
	250 to 328 mil crab	89.88*** (24.03)	chi2(1)=12.29*** p=0.0005
oysters	3,300 to 5,500 tons	81.69*** (21.14)	
	3,300 to 10,000 tons	61.27*** (23.25)	chi2(1)=0.97 p=0.3250
lakes	2,900 to 3,300 lakes	119.60*** (22.05)	
	2,900 to 3,850 lakes	152.20*** (26.22)	chi2(1)=2.74* p=0.0976

Testing for Preference Inconsistencies

A 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 simply revisiting their decision rule as they progress through the series of choice questions (Hensher 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 where interaction terms are included between the environmental and cost attributes with dummy variables denoting each choice occasion *t = 1, 2,* or *3*. The results are presented in table 24 below. There is no clear trend in the marginal utility estimates across the various choice occasions. Statistical tests fail to reject the null hypotheses that the parameter estimates for each attribute are statistically equal across the three choice questions. It appears that respondents' stated decisions are consistent across the choice questions, thus lending support to the notion that these stated choices are based on stable underlying preferences.

Table22 RUM Regressions: Testing for Consistent Preferences

	Question 1	Question 2	Question 3	H_0: Coefs equal
clarity	0.3043	0.4110*	0.2244	chi2(2)=0.46
	(0.2101)	(0.2291)	(0.1990)	p=0.7934
bass	0.0810***	0.0605*	0.0031	chi2(2)=4.12
	(0.0283)	(0.0327)	(0.0289)	p=0.0.1274
crab	0.0129***	0.0067	0.0062	chi2(2)=2.43
	(0.0036)	(0.0049)	(0.0038)	p=0.2969
oysters	0.0001***	0.0000	0.0001*	chi2(2)=2.63
	(0.0001)	(0.0000)	(0.0000)	p=0.2679
lakes	0.0012***	0.0014***	0.0013***	chi2(2)=0.35
	(0.0003)	(0.0004)	(0.0003)	p=0.8398
cost	-0.0094***	-0.0079***	-0.0067***	chi2(2)=2.15
	(0.0014)	(0.0011)	(0.0012)	p=0.3422

Testing for Relationships Predicted by Economic Theory

In this next subsection variants of the base model are estimated and used to examine whether respondents' stated choices exhibit behaviors predicted by economic theory, namely decreasing marginal utility of income and diminishing marginal benefit from increases in the environmental attributes. In table 25 the *cost* coefficient is allowed to vary across income categories by interacting *cost* with a dummy variable denoting each category. This relationship is clearly non-monotonic, but considering the entire income range there is some weak evidence of a decreasing marginal utility of income.

Table 23 RUM Results: Examining for Decreasing Marginal Utility of Income

VARIABLES	Income Category Interactions
clarity	1.930e-01
	(0.126)
bass	2.891e-02*
	(0.017)
crab	7.394e-03***
	(0.002)
oysters	4.038e-05
	(0.000)
lakes	1.101e-03***
	(0.000)
cost × 1(income 0-25k)	-5.605e-03***
	(0.002)
cost × 1(income 25-50k)	-5.788e-03***
	(0.001)
cost × 1(income 50-75k)	-9.297e-03***
	(0.002)
cost × 1(income 75-100k)	-7.319e-03***
	(0.001)
cost × 1(income 100-150k)	-8.607e-03***
	(0.002)
cost × 1(income 150-200k)	-5.391e-03***
	(0.002)
cost × 1(income 200k+)	-2.878e-03*
	(0.002)
status quo (mean)	-1.808e+00***
	(0.374)
status quo (std dev)	4.158e+00***
	(0.481)

Similar to the above assessment, economic theory suggests that utility is increasing in the environmental attributes but at a decreasing rate. The presence of nonlinearities in marginal utility is examined by adding squared terms to the base model. The base model results are shown in table 26 (column 1) for comparison. Column 2 includes squared terms for each of the environmental attributes: *clarity, bass, crab, oysters,* and *lake.* These squared terms are statistically insignificant for the first three attributes, both individual and jointly.[14] The sign and significance of the main and squared terms for oysters and watershed lakes, however, do exhibit evidence of diminishing marginal utility.

Table 24 RUM Results: Examining for Decreasing Marginal Utility of Environmental Attributes

VARIABLES	Linear Only	Quadratic
clarity	2.266e-01*	-2.719e+00
	(0.126)	(2.133)
clarity^2		3.939e-01
		(0.281)
bass	4.052e-02**	-2.106e-01
	(0.017)	(0.234)
bass^2		4.417e-03
		(0.004)
crab	8.755e-03***	-3.894e-02
	(0.002)	(0.049)
crab^2		8.306e-05
		(0.000)
oysters	5.704e-05**	7.236e-04***
	(0.000)	(0.000)
oysters^2		-4.899e-08***
		(0.000)
lakes	1.210e-03***	1.468e-02***
	(0.000)	(0.004)
lakes^2		-1.988e-06***
		(0.000)
cost	-7.400e-03***	-7.877e-03***
	(0.001)	(0.001)
status quo (mean)	-1.749e+00***	-1.630e+00***
	(0.358)	(0.374)
status quo (std dev)	4.127e+00***	4.319e+00***
	(0.473)	(0.500)

[14] Test for joint significance yielded a chi2(3)=4.88, p=0.1808.

8.2 Sensitivity Analyses

Sample Screening Criteria

As discussed in section 6.3, some respondents were identified as exhibiting protest, warm glow, or hypothetical biases, and were eliminated from the sample used in the main analysis. Such behaviors were identified based on responses to both debriefing questions and the choice questions themselves (see section 6.3 for details).

In tables 27 and 28 we present the household WTP for the improvements projected to occur under the TMDL under several alternative screening criteria in order to demonstrate the sensitivity of the WTP estimates to the screening criteria. These estimates are derived from the base model (Model 1), which only includes the environmental and cost attributes, and the random and normally distributed status quo constant (SQC). Going from top to bottom, the stringency for identifying protest responses increases, implying more respondents are screened out of the sample. Respondents identified as protesting the stated choice exercise are those who identified themselves as being against government regulations and not believing they should have to pay any amount to improve watershed lakes and the Chesapeake Bay. As such, one might expect such protests would bias WTP downward. Going from left to right, the criteria for identifying hypothetical bias and warm-glow become more stringent. One may expect such behaviors would bias WTP upward. In identifying the most appropriate sample for estimating the RUMs and WTP, it is important to treat both forms of biases symmetrically, as shown by the cells on the diagonals in tables 27 and28.

The most stringent screening criteria, and that used in the main analysis, corresponds to the cell in the third row, third column. It is re-assuring that the household WTP estimates in this sample are robust and fairly close to the estimates obtained under alternative assumptions. In fact, the estimates for all alternative criteria are within $12 or less of each other.

Table 25 Household WTP under Alternative Sample Screening Assumptions (Base Model): Based on Fish and Shellfish Projections from Expert Elicitation (2014 dollars))

	Unscreened	**Warm glow & Hypothetical Bias 1**	**Warm glow & Hypothetical Bias 2**
Unscreened	135.419*** (22.514) n=5,724	131.430*** (22.225) n=5,670	136.396*** (22.936) n=5,259
Protest 1	135.615*** (22.450) n=5,148	131.878*** (22.102) n=5,094	136.873*** (22.837) n=4,791
Protest 2	142.473*** (23.220) n=4,623	138.116*** (22.816) n=4,569	143.790*** (23.632) n=4,311

Table 26 Household WTP under Alternative Sample Screening Assumptions (Base Model): Based on Fish and Shellfish Projections from Multi-species Model (2014 dollars)

	Unscreened	**Warm glow & Hypothetical Bias 1**	**Warm glow & Hypothetical Bias 2**
Unscreened	86.057*** (15.131) n=5,724	83.424*** (14.969) n=5,670	88.095*** (15.540) n=5,259
Protest 1	86.166*** (15.080) n=5,148	83.387*** (14.870) n=5,094	88.310*** (15.453) n=4,791
Protest 2	87.339*** (15.442) n=4,623	84.394*** (15.225) n=4,569	89.591*** (15.806) n=4,311

Consideration of Omitted Variables

A general concern with stated choice experiments, such those posed in this survey, is that respondents may be thinking of other factors (or omitted variables) when answering the choice questions. If respondents correlate their perceptions of changes in such factors with one of the described attributes, then an omitted variable bias could potentially arise. Although such considerations could be perfectly valid in reality, how these factors change under each alternative was not clearly described to respondents, and in the absence of a clearly and quantitatively described mapping, respondents may assert their own beliefs, which could be widely unfounded and are unknown to the researcher (Boyd and Krupnick, 2013; Johnston et al., 2013). In this section such biases are closely examined, and the sensitivity of the WTP estimates assessed.

The base model (Model 1) is re-estimated with 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 stated they 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. Similarly, a series of interaction terms between *bass*, *crab*, and *oysters* with a dummy variable *food*, which denotes respondents that 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 regression are presented in table 29 with the base model results for comparison.

The coefficients corresponding to the interaction terms with *food* are statistically insignificant, both individually and jointly, suggesting that respondents who stated they considered the price and quality of seafood did not have systematically different preferences for the bass, crab, or oyster attributes than the rest of the sample.[15] The coefficients on *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.

Table 27 RUM Regressions with Interaction Terms denoting Respondents Who Considered Omitted Factors

VARIABLES	Model 1	Model 1 with Omitted Variable Interactions
clarity	2.266e-01*	2.315e-01*
	(0.126)	(0.125)
bass	4.052e-02**	2.056e-02
	(0.017)	(0.021)
crab	8.755e-03***	5.509e-03*
	(0.002)	(0.003)
oysters	5.704e-05**	3.311e-05
	(0.000)	(0.000)
lakes	1.210e-03***	7.410e-04***
	(0.000)	(0.000)
bass × food	4.127e+00***	4.230e-02
	(0.473)	(0.031)
crab × food	2.266e-01*	6.951e-03
	(0.126)	(0.004)
oysters × food	4.052e-02**	5.786e-05
	(0.017)	(0.000)
lakes × not watershed lakes	8.755e-03***	8.148e-04**
	(0.002)	(0.000)
cost	-7.400e-03***	-7.460e-03***
	(0.001)	(0.001)
status quo (mean)	8.755e-03***	-1.713e+00***
	(0.002)	(0.346)
status quo (std dev)	5.704e-05**	3.876e+00***
	(0.000)	(0.450)

Whether such influences reflect preference heterogeneity or omitted variable bias, they can be controlled for econometrically, and can be excluded from the resulting WTP estimates (by simply disregarding the coefficients corresponding to the *food* and *not watershed lakes* interaction terms). Table 30 shows household WTP estimates from the corresponding models in table 29. The total WTP

[15] Chi square tests fail to reject the null hypotheses that the coefficients corresponding to all three interaction terms with *food* are jointly equal to zero, chi2(3)=5.32 (p=0.1500).

estimates are presented in Table 31. Results show that controlling for differences in WTP between people who said they were considering factors other than the attributes listed in the choice questions reduces benefit estimates by about 35%.

Table 28 Comparing Annual Household WTP Estimates across Models Considering Omitted Factors

	Model 1	Model 1 with Interactions
High fish/shellfish projections	143.79*** (23.63)	91.91*** (26.80)
Low fish/shellfish projections	89.59*** (15.81)	58.76*** (18.25)

Table 29 Comparing Total Annual WTP Estimates across Models Considering Omitted Factors (billions of 2014 dollars)

Non-respondents' WTP	**Fish/shellfish**	Model 1	Model 1 with Interactions
Same as Respondents'	High	6.378*** (1.048)	4.077*** (1.189)
	Low	3.974*** (0.701)	2.606*** (0.809)
Zero	High	1.977*** (0.325)	1.264*** (0.369)
	Low	1.232*** (0.217)	0.808*** (0.251)

Comparison of Constant, Declining and Improving Baseline Versions

In addition to the constant baseline version of the survey we also administered two other versions in which respondents were told conditions would decline or improve in the future relative to current levels in the absence of additional actions. The declining baseline version of the survey was administered to the same geographic strata and with the same sampling intensity as the constant baseline version. The improving baseline version, however, was only administered in the Bay States stratum. Because the study areas were different for the declining and improving baseline surveys we perform separate comparisons to the constant baseline results.

On the declining baseline version of the survey the status quo option showed attribute levels below current levels with the attribute levels in the policy cases at or above current levels. Here we compare total WTP estimates using data collected with the two versions of the survey. Because the status quo option was the only scenario with declining attribute levels the alternative specific constant indicating the status quo option (SQC) is highly correlated with the attribute levels and must be dropped from the declining baseline model to estimate the coefficients and their standard errors. Table 32 reports the estimates form the base model with and without the SQC and the declining baseline model.

***Table** 30 Constant-Declining Baseline Comparison: Regression Results*

VARIABLES	Constant Baseline with SQC	Constant Baseline without SQC	Declining Baseline without SQC
clarity	2.266e-01*	1.833e-01**	3.767e-01***
	(0.126)	(0.081)	(0.069)
bass	4.052e-02**	2.412e-02**	1.484e-02
	(0.017)	(0.011)	(0.010)
crab	8.755e-03***	7.291e-03***	4.349e-03***
	(0.002)	(0.002)	(0.001)
oysters	5.704e-05**	3.195e-05*	-6.219e-06
	(0.000)	(0.000)	(0.000)
lakes	1.210e-03***	9.092e-04***	5.398e-04***
	(0.000)	(0.000)	(0.000)
cost	-7.400e-03***	-4.837e-03***	-4.688e-03***
	(0.001)	(0.000)	(0.000)
status quo (mean)	-1.749e+00***		
	(0.358)		
status quo (std dev)	4.127e+00***		
	-0.473		

The regression results of table 32 are used to estimate WTP for the changes in the attributes expected under the TMDL and reported in table 33. It is important to note that the WTP estimates in table 33 are all based on the same *change in attribute levels* and not the same policy outcome attribute levels. Applying the same policy outcome would mean that the declining baseline WTP would be calculated using larger attribute improvements than the constant baseline estimates.

Given that we are estimating the utility function with a linear approximation, if marginal utility is diminishing, one would expect the utility function estimated with declining baseline data to have a steeper slope than one estimated with constant baseline data. This would result in a larger total WTP in the declining baseline case for the same attribute improvements. Table 33 shows that, at least nominally, this does not appear to be the case. A t-test of means, however, fails to reject the null hypothesis that the declining baseline WTP is equal to either WTP estimated with the constant baseline data.

Table 31 Constant-Declining Baseline Comparison: Representative Household Annual WTP (2014 dollars)

Fish/shellfish projections	Constant Baseline with SQC	Constant Baseline without SQC	Declining Baseline without SQC
Low	89.59*** [58.61 - 120.57]	103.81*** [75.92 - 131.69]	84.23*** [65.71 - 102.77]
High	143.79*** [97.47 - 190.11]	169.71*** [132.63 - 206.80]	121.99*** [99.16 - 144.84]

Since the improving baseline version of the survey was only administered in the Bay States stratum we compare results across all three baseline versions of the survey but omit observations from the other two strata. The improving baseline version of the survey showed respondents a status quo option for which attribute improved moderately in the absence of further action. Under the policy choices all attributes improved by a larger margin than the baseline. Again, because the status quo option was the only choice with those attribute levels, the SQC must be omitted from the regression because of severe multicollinearity. Table 34 presents regression results from the base model across all three baseline versions using the Bay States observations only.

Table 32 Baseline Comparison Regression Results: Bay States Only

VARIABLES	Constant Baseline with SQC	Constant Baseline without SQC	Declining Baseline without SQC	Improving Baseline without SQC
clarity	2.674e-01	2.680e-01**	4.099e-01***	4.776e-01***
	(0.178)	(0.115)	(0.117)	(0.137)
bass	2.691e-02	1.313e-02	2.803e-02*	-2.508e-03
	(0.022)	(0.014)	(0.016)	(0.018)
crab	8.025e-04	3.251e-03	2.280e-03	4.909e-03**
	(0.003)	(0.002)	(0.002)	(0.002)
oysters	5.728e-05	5.843e-05**	2.212e-05	-7.758e-06
	(0.000)	(0.000)	(0.000)	(0.000)
lakes	8.554e-04***	7.501e-04***	4.676e-04***	4.905e-04*
	(0.000)	(0.000)	(0.000)	(0.000)
cost	-7.656e-03***	-4.453e-03***	-4.591e-03***	-3.971e-03***
	(0.001)	(0.001)	(0.001)	(0.001)
status quo (mean)	-2.149e+00***			
	(0.444)			
status quo (std dev)	4.040e+00***			
	(0.678)			

The regressions presented in table 34 are used to estimate the household WTP estimates in table 35. Again, it is the policy changes in attribute levels and not the final attribute levels resulting from the policy that are used to find WTP. If utility is diminishing in the environmental attributes then one would expect the WTP estimated with the linear approximation of the utility function to diminish as well as baseline condition improve. Table 35 shows the reverse relationship. WTP for the same improvement in environmental attributes is highest when estimated with improving baseline data, followed by the constant and declining baseline results. Similar to the declining baseline comparison, however, a t-test of means fails to reject the null hypothesis that the improving and declining baseline WTP estimates are the same as the constant baseline WTP when estimated with Bay States data.

Table 33 Baseline Comparison: Representative Household Annual WTP, Bay States only (2014 dollars)

Fish/Shellfish Projections	Constant Baseline with SQC	Constant Baseline without SQC	Declining Baseline without SQC	Improving Baseline without SQC
Low	64.799*** [24.88 - 104.71]	100.842*** [59.64 - 142.04]	81.495*** [51.90 - 111.08]	101.890*** [44.64 - 159.13]
High	75.417** [17.63 - 133.20]	138.455*** [85.93 - 190.97]	107.829*** [68.80 - 146.86]	148.606*** [90.58 - 206.62]

From the results of a number of validity tests and sensitivity analyses in this section we conclude that the WTP estimates performed well under tests of internal and external scope and showed reasonable sensitivity to attribute improvements of different magnitudes. Results of theoretical validity tests regarding diminishing marginal utility of income and environmental goods were mixed. WTP estimates were very stable across various screening criteria implying that even responses flagged by debriefing questions designed to catch certain systematic biases were fairly well-behaved. There is evidence of omitted variables bias with WTP estimates falling by about 35% when such considerations are controlled for econometrically, showing the importance of including debriefing questions on SP surveys to probe those issues. Finally, the WTP estimates estimated from different baseline versions of the survey were not statistically different from each other but nominally showed the reverse relationship one would expect if utility is diminishing in the provision of the environmental attributes.

9. Summary

The Clean Water Act (CWA) directs EPA to coordinate Federal and State efforts to improve water quality in the Chesapeake Bay. In 2009, Executive Order (E.O.) 13508 reemphasized this mandate, directing EPA to define the next generation of tools and actions to restore water quality in the Bay and describe the changes to be made to regulations, programs, and policies to implement these actions. In response, EPA has developed an assessment of the costs and benefits of meeting established pollution budgets, called Total Maximum Daily Loads (TMDLs), of nitrogen, phosphorus, and sediment for the Chesapeake Bay. The SP study presented in this report is a key component of that benefit cost analysis.

A suite of hydrological and ecological models are used to predict how five environmental attributes would change over time under the TMDLs and baseline conditions. Model predictions and valuation survey data are combined to estimate the total economic benefit of the management practices required to meet the TMDLs. 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.

Using a choice experiment format, EPA is able to estimate the marginal willingness to pay for changes in the Bay populations of blue crabs, oysters and striped bass, water clarity in the Chesapeake Bay and changes in algae levels in lakes in the watershed. The survey was administered to a stratified sample in 17 eastern states and the District of Columbia allowing EPA to extrapolate household WTP estimates to the study region

Results show that a representative household in the study area are willing to pay between $82 and $167 per year for water quality improvements expected as a result of the TMDLs. The variation in per household WTP reflects different projections for fish and shellfish populations. As expected, users are willing to pay more for improvements in environmental attributes than non-users. The survey data perform well in internal and external tests of scope and several other tests of theoretical validity.

According to the 2010 decennial Census, there are a total of 44,353,441 households in the 17 states and District of Columbia. Applying the representative household WTP estimates to this population of households results in a range of total WTP of $1.134 billion to $5.901 billion. The lower bound applies the lower fish and shellfish projections and assumes the proportion of the population corresponding to survey non-respondents have a zero WTP for the environmental improvements expected under the TMDLs. The upper bound applies the larger increases in fish and shellfish populations and assumes the population of the study area has the same willingness to pay as respondents.

These estimates should be interpreted with a clear understanding of the scope of benefits described in the survey. Respondents were asked to value changes in 5 attributes due to changes in phosphorus, nutrients and sediment in the Bay. Results reflect values elicited from both users of the Bay or lakes in the Watershed (i.e., respondents who have engaged in recreation activities in these areas in the last 5 years) and non-users. The WTP of non-users, while lower than that of users at the household level, constitutes a large proportion of the total benefits because an estimated 71%[16] of the households in the study area are non-users of the Chesapeake Bay or Watershed lakes. Only stated preference methods like the study presented here can capture this component total economic value.

References

Alpızar, Francisco, Fredrik Carlsson, Peter Martinsson (2001) “Using choice experiments for non-market valuation,” *Economic Issues* 8, 83–110.

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

Bennett, J., Adamowicz, W., 2001. Some fundamentals of environmental choice modelling. In: Bennett, J., Blamey, R. (Eds.), The Choice Modelling Approach to Environmental Valuation. Edward Elgar, Cheltenham.

Bennett, Jeffrey W, Russell K Blamey (2001), *The Choice Modelling Approach to Environmental Valuation,* Edward Elgar, Cheltenham, UK.

[16] Of the 264 respondents to the NRFU 188 had not visited either a lake in eth Watershed or the Bay for recreation in the past 5 years.

Bergstrom, John C, Laura O Taylor (2006). “Using meta-analysis for benefits transfer: theory and practice,” *Ecological Economics* 2, 351–60.

Bergstrom, John C., John R. Stoll, and Alan Randall. "The impact of information on environmental commodity valuation decisions." *American Journal of Agricultural Economics* 72, no. 3 (1990): 614-621.

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, Sept). *The definition and choice of environmental commodities for nonmarket valuation*. RFF DP 09-35, Resources for the Future discussion paper.

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.

Boyle, Kevin J. "Commodity specification and the framing of contingent-valuation questions." Land Economics (1989): 57-63.

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.

Carson, Richard T, Mikolaj Czajkowski (2014) “The discrete choice experiment approach to environmental contingent valuation,” in: Hess, Stephane, Andrew Daly, (Eds.), *Handbook of Choice Modelling.*

Carson, Richard T, Robert C Mitchell (1993) “The value of clean water: The public’s willingness to pay for boatable, fishable, and swimmable quality water,” *Water Resources Research*, 29, 2445–54.

Carson, Richard T. (2012) “Contingent valuation: a practical alternative when prices aren’t available,” *Journal of Economic Perspectives,* 26(4), 27-42.

Carson, Richard, JJ Louviere (2011), “A common nomenclature for stated preference elicitation approaches,” *Environmental & Resource Economics*, 49, 539-559.

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

Desvousges, William H, V Kerry Smith, Ann Fisher (1987) “Option price estimates for water quality improvements: A contingent valuation study for the Monongahela River,” *Journal of Environmental Economics and Management*, 14, 248–67.

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

Groves, R. M., Couper, M. P., Presser, S., Singer, E., Tourangeau, R., Acosta, G. P., et al. (2006) Experiments in producing nonresponse bias. Public Opinion Quarterly, 70(5), 720-736.

Halpern, S. D., D. A. Asch, A. Shaked, P. G. Stock, and E. Blumberg. "Standard definitions: Final dispositions of case codes and outcome rates for surveys." *AMERICAN JOURNAL OF TRANSPLANTATION* 5, no. 6 (2005): 1319-1325.

Hausman, Jerry (2012) "Contingent Valuation: From Dubious to Hopeless," *Journal of Economic Perspectives,* 26, 43-56.

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, Stephane, David A. Hensher, and Andrew Daly (2012), "Not bored yet – Revisiting respondent fatigue in stated choice experiments," Transportation Research Part A: Policy and Practice, 46(3), 626-644.

Hicks, Robert, Ivar Strand (2000), "The extent of information: its relevance for random utility models," *Land Economics*, 76, 374-385.

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.

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.

Iannacchione, V. G. (2011). The changing role of address-based sampling in survey research. *Public Opinion Quarterly, 75*(3), 556-575.

Jeon, Yonsik, Joseph A Herriges, Catherine L Kling, John A Downing (2005), "The role of water quality perceptions in modeling lake recreation demand," in Iowa State University Department of Economics Working Paper Series, Working Paper #05032. Ames: Iowa State University.

Johnston, R.J., Opaluch, J.J., Mazzotta, M.J., and Magnusson G. (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, 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, PP.

Johnston, Robert J, Randall S Rosenberger (2010), "Methods, trends and controversies in contemporary benefit transfer," *Journal of Economic Surveys,* 24, 479–510.

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.

Kling, Catherine L, Daniel J Phaneuf, Jinhua Zhau (2012), "From Exxon to BP: Has Some Number Become Better than No Number," *Journal of Economic Perspectives,* 26, 3-26.

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.

Krupnick, Alan (1988), "Reducing bay nutrients: An economic perspective," *Maryland Law Review,* 47, 453–480.

Leggett, Christopher G, Nancy E Bockstael (2000) "Evidence of the effects of water quality on residential land prices," *Journal of Environmental Economics and Management*, 39, 121–44.

Lindsey, Greg H, Robert G Paterson, Michael L Luger (1995), "Using contingent valuation in environmental planning," *Journal of the American Planning Association*, 61, 252–62.

Link, M.W. Battaglia, M.P., Frankel, M.R., Osborn, L., and Mokdad, A. H. (2008). A comparison of address-based sampling (ABS) versus random-digit dialing (RDD) for general population surveys. *Public Opinion Quarterly, 72*(1), 6-27.

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

Lipton, Douglas W, Robert L Hicks (1999), "Boat location choice: the role of boating quality and excise taxes," *Coastal Management*, 27, 81-90

Lipton, Douglas W, Robert L Hicks (2003), "The cost of stress: low dissolved oxygen and economic benefits of recreational striped bass (Morone saxatilis) fishing in the Patuxent River," *Estuaries, 26*, 310-315.

Massey, D Matthew, Steven C Newbold, Brad Gentner (2006), "Valuing water quality changes using a bioeconomic model of a coastal recreational fishery," *Journal of Environmental Economics and Management,* 52, 482–500.

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.

Morgan, Cynthia, Nicole Owens (2001), "Benefits of water quality policies: the Chesapeake Bay," *Ecological Economics*, 39, 271-284.

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.

Poor, P Joan, Keri L Pessagno, Robert W Paul (2007), "Exploring the hedonic value of ambient water quality: A local watershed-based study," *Ecological Economics, 60*, 797-806.

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

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.

Smith, V Kerry, William H Desvousges, and Ann Fisher (1986), "A comparison of direct and indirect methods for estimating environmental benefits," *American Journal of Agricultural Economics*, 68, 280–90.

USEPA (U.S. Environmental Protection Agency). 2010. Chesapeake Bay Phase 5.3 Community Watershed Model. EPA 903S10002 - CBP/TRS-303-10. U.S. Environmental Protection Agency, Chesapeake Bay Program Office, Annapolis MD. December 2010.

Van Houtven, George (2009), "Changes in ecosystem services associated with alternative levels of ecological indicators," In *Risk and Exposure Assessment for Review of the Secondary National Ambient Air Quality Standards for Oxides of Nitrogen and Oxides of Sulfur*. Research Triangle Park, NC: RTI International.

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.

Walsh, Patrick, William Wheeler (2013), "Water quality index aggregation and cost-benefit analysis," *Journal of Benefit Cost Analysis*, 4, 81-106.

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, QLD, pp. 19.

1 Appendix