NCEE Working Paper

Muddying the Water? An Analysis of Non-Constant Baselines in Stated Preference Surveys

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

Working Paper 18-02
May, 2018

U.S. Environmental Protection Agency
National Center for Environmental Economics
https://www.epa.gov/environmental-economics

NATIONAL CENTER FOR ENVIRONMENTAL ECONOMICS

Muddying the Water? An Analysis of Non-Constant Baselines in Stated Preference Surveys

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

Abstract: Defining baseline conditions is a key component of regulatory benefit-cost analysis. Most stated preference studies assume that the current state of the world in the absence of additional policy action remains constant. In the time that passes while a regulation is evaluated, implemented, and produces the intended environmental impacts, however, this is unlikely to be the case. To address this largely unexplored area of nonmarket valuation, we administer a stated preference survey using a three-way split sample design. Respondents are either told future baseline conditions would remain constant, decline, or improve without additional policy interventions. While we find some evidence to support predictions of the standard theoretical model, we also find that behavioral and emotional reactions to the non-constant baseline scenarios muddy the waters, introducing some countervailing factors. These results have implications for the design and use of stated preference results in benefit-cost analysis.

JEL Codes: Q51, Q53

Key words: baseline, benefit-cost analysis, Chesapeake Bay, nonmarket valuation, stated preference survey

Muddying the Water? An Analysis of Non-Constant Baselines in Stated Preference Surveys

Kelly B. Maguire,* Chris Moore,* Dennis Guignet,* Chris Dockins,* and Nathalie B. Simon[1]

I. INTRODUCTION

Defining the baseline conditions is a key component of regulatory benefit-cost analysis. The baseline is a reference point or counterfactual that reflects conditions without the policy under evaluation in place. All benefits and costs are then measured relative to this baseline. The U.S. Environmental Protection Agency's (EPA) *Guidelines for Preparing Benefit Cost Analyses* (USEPA, 2010a) devotes an entire chapter to the topic, yet few valuation studies examine the implications of alternative baseline conditions on both theoretical and empirical assessments of benefits of policy interventions. Evaluating a policy that is expected to only be fully effective in future years raises some challenges in specifying baseline conditions. There is uncertainty about how population growth and land use changes will affect environmental quality. In addition, new technology, environmental practices, and the impact of existing policies can affect the baseline, which in turn can affect the scale of the expected environmental improvements and possibly the marginal willingness to pay for the improvements. It is not always feasible to resolve these uncertainties prior to an intended policy analysis, and while multiple treatments with alternative baselines may be presented, most benefit estimates used to monetize changes in environmental conditions are based on studies that elicit WTP for an improvement relative to current conditions or, what we refer to as a constant baseline.

*National Center for Environmental Economics, US EPA
[1] Corresponding author, National Center for Environmental Economics, US EPA, Mail Code 1809T, 1200 Pennsylvania Avenue, NW, Washington, DC 20460. Phone: 202-566-2347. Email: simon.nathalie@epa.gov.

Indeed, many published studies present scenarios using a single, constant baseline with only a few exploring the effects on willingness to pay of alternative baseline treatments (Abt Associates 2016; Banzhaf *et al*. 2006; Soto Montes de Oca and Bateman 2006; Lew *et al*. 2010). Using a single, constant baseline satisfies a number of desirable survey design objectives, including: parsimony (keeps the survey shorter); familiarity (respondents are more likely to have some knowledge of current conditions); cognitive burden (respondents only have to compare the policy scenario to current conditions rather than to alternative future states of the world); and scenario rejection (respondents may be less likely to believe a baseline description that differs from current conditions).[2] However, our interest in this topic is not merely academic. If future conditions differ from the current status quo and WTP estimates are sensitive to alternative baseline characterizations, then there are implications for benefit transfer, on which agencies like the EPA often rely when preparing benefit-cost analyses. EPA's *Guidelines* recommend that suitable studies used in benefit transfer be as similar as possible to the policy case in their "baseline and extent of environmental changes" (USEPA 2010a).

The stated preference literature offers few insights on the sensitivity of WTP to alternative baselines. Lew *et al.* (2010), for instance, consider three alternative baseline scenarios for Stellar sea lion populations when eliciting marginal WTP for protection programs, finding that WTP decreases as the forecasted future baseline population improves. They also find evidence of diminishing marginal utility for population increases that exceed current levels. Banzhaf *et al*. (2006) employ two baselines (one constant and one declining) to "bracket the range of

[2] Johnston, *et al*. 2017 provide a set of best practices to apply and consider when conducting stated preference surveys.

uncertainty in the science" and the expected environmental status in the absence of interventions in their valuation survey of programs in the Adirondack Park. While they find higher WTP under declining baseline conditions, results were confounded by differences in the overall level of improvement in the attributes. Using a dichotomous choice survey, Soto Montes de Oca and Bateman (2006) consider the effects of different baseline conditions on WTP for the provision of water services in Mexico City. They find that households with better baseline water quality and fewer disruptions in service had lower WTP for improvements than households with lower baseline quality and service levels.

These studies underscore the importance of alternative baselines in stated preference surveys. Our study offers additional evidence while explicitly accounting for potential biases that may influence the results. We report results from a stated preference survey for improvements in ecological conditions in the Chesapeake Bay and lakes in the Bay Watershed under three different assumptions regarding future baseline conditions. Specifically, we use a discrete choice experiment and three-way split-sample design to empirically test differences in marginal and household willingness to pay for clean-up programs under a constant baseline (i.e., current ecological conditions are expected to remain the same in future years); an improving baseline (i.e., conditions will improve in the future due to existing programs, although not as much as they could if new programs were implemented); and a declining baseline (i.e., conditions will worsen in future years without the implementation of new programs due to factors like population growth and land use change). Additionally, our split sample design provides a unique opportunity for a test of external scope. Specifically, we estimate household WTP for a common policy goal above the improving baseline to avoid extrapolating outside of the observed attribute space for all three baseline samples.

Because few previous stated preference studies use baseline scenarios that differ from current conditions there are several practical issues to address. Asking survey respondents to consider improvements relative to baseline conditions that are different from current conditions, and could be unfamiliar to respondents, increases the cognitive burden of the choice task. Greater cognitive burden increases the possibility that respondents could apply simplifying heuristics for decision making or dismiss some of the tradeoffs between multiple attributes and cost. It is also conceivable that describing future conditions that are different from current conditions will exacerbate some of the challenges of stated preference methods, such as scenario rejection and strategic responses, which can otherwise be mitigated by careful survey design. We examine the data for evidence of these issues and compare results across baseline scenarios.

The remainder of the paper is organized as follows. Section II provides the theoretical foundation underlying our study and hypotheses we test. Section III presents our empirical model; we describe the survey instrument and data in Section IV. We present our results in Section V and conclusions in Section VI.

II. THEORETICAL FOUNDATION

The typical exposition of welfare economics for non-market valuation begins with the direct utility function that is not only a function of private goods, *x*, but also public goods, *q*.

$$u = f\left(x, q\right) \tag{1}$$

Individuals choose their level of consumption of *x* but the provision of *q* is exogenous. Take, for example, a person's choice over boating trips. Consumption of the private good, the amount of time on the water, is chosen by the individual, but the water quality, the condition of the boat launch, and the quality of the views are public goods that the boater takes as given when making the consumption decision. Individuals will maximize their utility subject to income, *Y,* yielding the indirect utility function,

$$V\left(q, p, Y\right) = \max_{x} \left\{ u\left(q, x, Y\right) \mid p \cdot x \leq Y \right\}, \tag{2}$$

where *p* is a vector of prices for the private goods.[3] Society's willingness to pay (WTP) for improvements in public goods, such as environmental quality, is a valid measure of the welfare gains derived from those improvements and can be defined implicitly by:

$$V\left(q_0 + \Delta, p, Y - WTP\right) = V\left(q_0, p, Y\right) \tag{3}$$

in which Δ is a vector of improvements to the baseline level of environmental quality, q_0. This is a compensating measure of the welfare change where the reference is initial utility. Most non-market valuation studies are concerned with the relationship between *WTP* and Δ or, in other words, how welfare gains change with the size of the environmental improvements. We broaden this conventional focus by also examining the relationship between *WTP* and q_0, or how the welfare gains are affected by baseline conditions.

The relationship between *WTP* and q_0 will depend largely on the shape of the utility function. Preferences that are strictly convex over *q* should result in marginal WTP that decreases as baseline conditions improve for a given Δ, while nonconvex preferences could result in marginal WTP that is alternately increasing and diminishing. Only if preferences are linear over the relevant range of *q* will baseline conditions have no impact on marginal WTP. The practical

[3] Consumption and prices of private goods (*x* and *p*, respectively) are not of primary interest in the later empirical analysis, and are thus represented as a composite numeraire good in the subsequent empirical model.

implication for policy analysis is that an accurate representation of baseline conditions when collecting stated preference data could be critical to estimating benefits correctly.

III. EMPIRICAL MODEL

Discrete choice experiments present respondents with alternative options, each describing a scenario with several different attributes (Alpizar *et al.,* 2001; Bennett and Adamowicz, 2001; Carson and Czajkowski, 2014). One of the attributes is the cost of each scenario, specifying some monetary amount a respondent must pay if that scenario is chosen. Respondents are asked to choose their preferred option from the available choices, where the levels of the attributes, including costs, vary across the scenarios. Each respondent's choice reflects their preferred trade-offs between attributes. By evaluating the choices made by respondents one can infer relative values and, by using the cost attribute, estimate marginal willingness to pay for a change in each attribute.

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

$$u_{ij} = v\left(\boldsymbol{q_j}, Y_i - C_j\right) + \varepsilon_{ij} = v_{ij} + \varepsilon_{ij}, \qquad (4)$$

where $v(\cdot)$ or v_{ij} is the deterministic component of utility, and is a function of a vector of attributes describing the level of environmental quality $(\boldsymbol{q_j})$, as well as numeraire consumption, $Y_i - C_j$ (income minus the cost of alternative *j*). Utility also depends on a stochastic component, ε_{ij}, that is not observable to the researcher. For choice question t, respondent *i* will choose alternative *j* if it yields the greatest utility over all other available alternatives,

$$v_{ijt} + \varepsilon_{ijt} > v_{ikt} + \varepsilon_{ikt}, \forall\, k \neq j\,. \qquad (5)$$

The literature offers no clear guidance regarding the choice of specific functional forms for $v(\cdot)$. In practice linear forms are often used (Johnston *et al.,* 2003), although some studies have applied more flexible forms to allow for nonlinearities over the attribute space (e.g., Cummings *et al.*, 1994). We adopt a linear-in-logs model, where the environmental attributes enter $v(\cdot)$ in natural log form. This allows us to capture diminishing marginal utility while preserving more degrees of freedom than a model with higher order effects.

We estimate separate models for each baseline scenario. The model specification includes an alternative specific constant identifying the status quo alternative in each choice question. This status quo constant (SQC) serves to test and, if needed, control for respondents' preferences for the status quo option, irrespective of the attribute improvements and cost. A positive SQC suggests respondents tend to favor the status quo, perhaps reflecting either protest responses or "cold feet" towards a policy option. In contrast, a negative SQC suggests respondents favor a policy option in general, regardless of the environmental improvements specified. Such behavior could be due to respondents considering omitted factors (i.e., improvements to aspects of the environment that are not described by the choice attributes), or a general warm-glow for doing something to help the environment. Conceptually, the status quo effect is part of the indirect utility function, possibly capturing impacts known to the respondent but not the researcher. At the same time, it could reveal biasing behaviors that should be omitted from welfare analysis (see Boxall *et al.* 2009). We discuss the implications of these alternative interpretations on our WTP estimates.

When calculating the probability that respondent *i* chooses alternative *j*, the un-interacted income term Y_i drops out and the utility function becomes:

$$u_{ijt} = \boldsymbol{\beta_i} ln(q_{ijt}) + \gamma C_{ijt} + SQC_i \cdot d_j + \varepsilon_{ijt}, \tag{6}$$

where γ is the negative of the marginal utility of income (i.e., $\gamma = -\psi$, where ψ is the marginal utility of income), SQC_i is the status quo constant, and d_j is an indicator variable equal to one if alternative *j* corresponds to the status quo, and zero otherwise. Notice that SQC_i and $\boldsymbol{\beta_i}$ are specified as random coefficients that vary for each respondent i. In our empirical application, we assume SQC_i and $\boldsymbol{\beta_i}$ follow a normal distribution, allowing for respondents who may react positively or negatively towards the status quo across the different baseline versions of the survey and preference heterogeneity with respect to the environmental attributes, respectively. The marginal utility of income, $-\gamma$, is assumed to be fixed to ensure the existence of the WTP distribution (Daly *et al.,* 2012). Assuming ε follows a type I extreme value (Gumbel) distribution allows us to analyze responses via a mixed logit model (McFadden and Train, 2000).

For notational ease, let $\boldsymbol{\theta_i}$ denote a parameter vector encompassing all random coefficients (SQC_i and $\boldsymbol{\beta_i}$,), and let $\boldsymbol{x}_{ijt}$ be a vector including both $ln(\boldsymbol{q}_{ijt})$ and d_j**.** The probability of observing respondent *i*'s choices over the *T*=3 scenarios offered in each choice set in the survey is calculated by solving the integral:

$$P_i = \int \prod_{t=1}^{3} \left\{ \frac{\boldsymbol{exp}(\boldsymbol{\theta_i x}_{ijt} + \gamma C_{ijt})}{\sum_{\boldsymbol{k=1}}^{3} \boldsymbol{exp}(\boldsymbol{\theta_i x}_{ikt} + \gamma C_{ikt})} \right\} \varphi(\boldsymbol{\theta}|\boldsymbol{b}, \boldsymbol{W}) d\boldsymbol{\theta}, \tag{7}$$

where $\varphi(\boldsymbol{\theta}|\boldsymbol{b}, \boldsymbol{W})$ is the normal density with mean vector ***b*** and covariance matrix ***W****.* The above integral has no analytical solution but can be approximated by simulation. The parameters ***b*** and ***W*** are found via maximum simulated likelihood providing population means for the utility parameters and an indication of heterogeneity in preferences for the choice attributes.

The vector of average marginal willingness to pay (MWTP) estimates for the environmental attributes is estimated as:

$$\boldsymbol{MWTP}(\overline{\boldsymbol{q}}) = \frac{\boldsymbol{\beta}}{-\gamma\overline{\boldsymbol{q}}}, \tag{8}$$

where $\overline{\boldsymbol{q}}$ denotes a reference level because utility is nonlinear over the attribute space.

Household willingness to pay (WTP) for an improvement in the environmental attribute vector from $\boldsymbol{q_0}$ to $\boldsymbol{q_1}$ is calculated following Holmes and Adamowicz (2003) as:

$$\boldsymbol{WTP}_{\boldsymbol{q_0} \to q_1} = \frac{\left(\boldsymbol{\beta} ln(\boldsymbol{q_1}) - \boldsymbol{\beta} ln(\boldsymbol{q_0})\right)}{-\gamma}. \tag{9}$$

Notice that in equation (9), the SQC is not included in the welfare calculations. While the SQC may capture valid welfare impacts such as those arising from omitted variables known to the respondent but not the researcher (Boxall *et al*. 2009), we follow standard practice and exclude it from the welfare calculations. To examine the impact this adjustment has on WTP estimates, we separately monetize the status quo effect by dividing the estimated impact on indirect utility by the marginal utility of income:

$$\boldsymbol{WTP}_{\boldsymbol{SQC}} = \frac{\boldsymbol{SQC}}{-\gamma}. \tag{10}$$

IV. THE CHESAPEAKE BAY SURVEY and DATA

We administered a stated preference survey to estimate WTP for attributes that would change because of nutrient and sediment loading reductions into the Chesapeake Bay. The survey was administered in 18 states in the eastern United States (U.S.), including those that have shoreline on the Bay and those that contain any part of the Chesapeake Bay Watershed. Each survey included three choice questions, where respondents chose the status quo option or one of two policy scenarios. Options were characterized by a set of environmental attributes in the year 2025 and household costs. Respondents were shown conditions today (i.e., levels of the attributes) and in 2025. With respect to new programs and improvements in the attribute levels, respondents were told the programs would be phased in over time and environmental conditions would improve before reaching long term conditions by 2025. The attributes included water clarity, populations of striped bass, crab, and oysters in the Bay, as well as the number of lakes in the broader Chesapeake Bay Watershed that have "low" algae levels.[4]

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 because of nutrient and sediment load reductions. The levels of the attributes are based on results of the Chesapeake Bay Watershed Model (USEPA, 2010b), an expert panel convened to predict effects on fish and shellfish populations (Massey *et al.*, 2017), and the Northeast Lakes Model (Moore *et al.*, 2011). Cost levels ensure adequate coverage of the WTP distribution and are based on focus group and pretest results.

The survey consisted of several sections to inform respondents about the Chesapeake Bay and programs to improve conditions in the Bay and lakes in the Watershed, followed by the choice questions, attitudinal responses, and demographic information. Table I provides the status quo levels for each baseline version, along with the levels of the policy options. Each randomly selected household from the sample frame was mailed a pre-notification letter followed by the survey with a cover letter, all bearing the EPA seal to increase response rates and better convey consequentiality. Households that did not return a survey within 4 weeks received a reminder post card and, eventually, a final reminder letter with a second copy of the survey booklet (Dillman, 2008). We conducted a pretest in late 2013 and the main survey in May 2014.

[4] "Low" algae lakes were defined as those having a lower than hypertrophic state. Details about the lakes and other environmental quality attributes are described by Moore *et al.*, (2018).

Because the surveys were nearly identical, we combine data from the pretest and main survey in the analysis. See Moore *et al.* (2015) for a full discussion of the survey development process, which included extensive focus groups and cognitive interviews. Table II shows the number of responses and response rates for the sample.

We utilized a stratified random sampling plan, where the survey was mailed to a random sample of households located within each of three geographic strata: Bay States, Watershed States, and Other East Coast States.[5] The survey instrument was mailed to 1,620 households in the pretest and 6,601 households in the main survey. All three baseline versions were mailed and allocated equally across the three geographic strata for the pre-test. In the main survey, the improving baseline was only implemented in the Bay States strata due to budget constraints. Table III provides demographic and attitudinal information and comparisons across the different baseline versions of the survey. The composition of respondents is relatively similar across baseline versions. The only significant difference we find is fewer Hispanics in the constant versus improving baseline samples, and fewer Blacks in the declining versus improving baseline samples.

Respondents to the improving baseline version are more likely to have heard of or visited the Bay and Watershed compared to the other two versions. Awareness of pollution in the Bay is similar across the three baseline versions. The significant differences in demographics and familiarity with the Bay between the improving baseline sample and the other two baseline versions is likely a result of the sample design. The improving baseline version was only

[5] The Bay States stratum consisted of all states adjacent to the Chesapeake Bay tidal waters (Maryland, Virginia, as well as the District of Columbia). The Watershed States included all states at least partially within the Chesapeake Bay Watershed, but not adjacent to the Bay (Delaware, New York, Pennsylvania, and West Virginia). The Other East Coast States stratum consisted of all other states within 100 miles of the US East Coast (Connecticut, Florida, Georgia, Maine, Massachusetts, New Hampshire, New Jersey, North Carolina, Rhode Island, South Carolina, and Vermont).

administered outside of the Bay States during the pretest resulting in a larger proportion of Bay States residents in the improving baseline sample. The constant and declining versions of the survey were assigned randomly, and mailed equally across the geographic strata for both the pretest and main survey. Despite the systematic difference in sampling, there is still value in comparing the improving baseline results to those from the constant and declining baseline versions of the survey. By using all data, we find a more robust set of results that indicate some behavioral responses across baselines, as discussed in the next section. Nonetheless, any comparison and interpretation with respect to the improving baseline must be caveated appropriately. We discuss the robustness of our findings at the end of the results section.

V. RESULTS

We perform four comparisons across baseline versions of the survey to fully assess the impact baseline conditions may have on responses to the choice questions.

Scenario Acceptance and Consequentiality

First, we compare measures of validity across baseline versions to address the practical question of whether the baseline affects the reliability of survey responses, perhaps due to scenario rejection or some other mechanism. Scenario rejection occurs when respondents fail to accept the choice scenario as presented. They may view the descriptions of the policy as unrealistic or object to the provision mechanism or payment vehicle. Scenario adjustment is a related respondent behavior that may ultimately have the opposite effect on responses. Rather than rejecting the scenario outright, respondents may substitute their own subjective beliefs about improvements under the policy or costs to their household and choose an option based on that set of information rather than what is presented in the survey (Cameron *et al.*, 2011).

We include two debriefing questions to probe for scenario rejection and adjustment. Each uses a Likert-scale response format with values from 1 to 5, from “Strongly Disagree” to “Strongly Agree.” The first statement is, “I voted as if my household would actually face the costs shown.” We interpret Disagree or Strongly Disagree responses as rejections or adjustments of the payment scenario. As shown in Table IV, we see the highest percentage of these responses (5.9 percent) in the constant baseline version. But the declining and improving baseline samples were not statistically different using a two-sample t-tests of proportions. The second question is, “I voted as if the programs would achieve the results shown.” Disagreeing with this statement is an indication that respondents rejected the scenario outright or possibly substituted their own subjective beliefs about the improvements that would occur under the provision mechanism. Again, the differences across baseline samples are not statistically significant.

An additional debriefing question probes on the consequentiality of the payment scenario. Using the same Likert-scale response format, the question prompt reads, “It is important to improve the waters of the Chesapeake Bay, no matter how high the cost.” Table IV, shows that large proportions of all three samples either agreed or strongly agreed with this statement, which calls into question whether these respondents realized the fiscal implications of the survey and were instead using their response to indicate general support for the programs. Further, there is a statistically significant difference between the responses of the constant and declining baseline samples to this prompt. When people are told that conditions are going to decline without policy intervention, they are more likely to disregard the stated costs of each program and support the policy. This is the first indication that baseline conditions described on the survey may affect the way people respond to the choice questions.

Mixed-logit comparison

The remaining comparisons are performed on a screened sample where we remove responses showing the strongest evidence of scenario rejection or adjustment.[6] Table V shows estimation results for each of the three baseline samples. In the constant baseline, the mean coefficients on all environmental attributes and cost are of the expected sign and are statistically significant. The SQC variable is negative and significant indicating a tendency to choose one of the policy options not explained by changes in the environmental attributes or cost. This holds, to varying degrees, across all three samples.

In the improving baseline sample, only Bay water clarity and the blue crab population are statistically significant. Although all mean coefficients on attributes exhibit the expected positive sign, striped bass populations, oyster abundance, and the number of low-algae Watershed lakes are not statistically significant. Yet, the cost coefficient and the SQC are statistically significant and negative. It may be that preferences toward additional improvements, above those already expected at no additional cost, are relatively weak. As such, further improvements do not have a significant impact on the likelihood of choosing a policy option with an additional cost.

The declining baseline results do not support this explanation, however. We again find clarity and one other attribute (striped bass in the declining baseline) are statistically significant. However, if satiation was the explanation for the marginal utilities being statistically equal to zero in the improving baseline sample, we would expect to see positive marginal utilities to all environmental attributes in the declining baseline results. Further, diminishing marginal utility

[6] See Moore *et al*. (2018) for a description of the sample screening criteria. The estimated coefficients, standard errors, and mean marginal WTP estimates using the full unscreened sample are presented in Appendix A, and are similar to those estimated here using the screened sample.

suggests that the mean coefficient estimates should be positive and of a higher magnitude compared to the constant baseline results. Since this is not the case, we explore other explanations. One possible explanation is that non-constant baselines increase the cognitive burden of the choice task, causing respondents to resort to simplifying heuristics when choosing among options. Rather than considering tradeoffs among all attributes respondents may focus on a subset they care about most, resulting in attribute non-attendance (Boxall *et al.*, 2009). Respondents may consider water clarity in the Bay as an overall indicator of aquatic ecosystem health, allowing them to attend less, or not at all, to the other Bay attributes when faced with the more complex baseline scenarios, even if they do have positive preferences for these amenities. Or, it may be that since water clarity appeared first in every choice question in all versions that respondents focused primarily on this attribute when answering more cognitively taxing questions involving non-constant baselines.

Another key difference between the mixed-logit results across the samples is the magnitude of the status quo constant in the declining baseline sample compared with the other two baseline samples. While we must use caution when comparing the parameter estimates across samples because of possible differences in the utility scale parameter (Swait and Louviere, 1993), the status quo constant in the declining baseline sample is twice the magnitude of the estimates from the other two samples, indicating a stronger tendency to choose a policy option that is not explained by the environmental improvements or cost. This could indicate an emotional response to worsening conditions (i.e., "something must be done"), while not carefully considering tradeoffs among the policy outcomes. Estimating the status quo constant and omitting it from WTP calculations we show later removes this potential effect from the welfare estimates. The resulting household WTP estimates are adjusted downward by $229 and $217 in the constant and improving baseline samples, respectively. The same adjustment in the declining baseline sample is more than double this amount, with a mean of $513.

Very few stated preference studies take the additional step of monetizing the tendency to vote for a policy, or alternatively the status quo option, that remains unexplained by other covariates. One notable exception is Boxall *et al.* (2009) who estimate a random status quo coefficient and find an average tendency to choose the status quo when the choice task is more complex, where complexity is measured as a change in multiple attributes, as opposed to just one attribute. As such, when the status quo effect is controlled for, WTP estimates are approximately $300 higher – an adjustment that is similar in magnitude, though of the opposite sign, to ours. The tendency to ignore trade-offs across attributes and options is a similar finding across the two studies.

Marginal Willingness to Pay

The differences in the mixed-logit model results in Table V are apparent in the marginal WTP estimates shown in Table VI as well. Each estimate of marginal WTP is generated using equation (5) and the baseline attribute levels from the corresponding version of the survey. Since clarity is the only attribute that is significant across all three samples, we restrict our inference about the shape of the utility function to clarity estimates. The mean marginal WTP for clarity in the declining baseline sample is greater than that of the constant baseline sample, as one would expect if preferences are convex. That trend is reversed, however, when baseline conditions continue to improve. Indeed, the marginal WTP for additional clarity is largest in the improving baseline sample – although it is statistically indistinguishable from the declining baseline estimate. These results suggest a utility function with greatest utility gains at low and high levels of clarity, but leveling off at current levels. Given indications of cognitive burden discussed earlier, however, respondents may have focused on clarity as an overall indicator of water quality when faced with non-constant baselines. Such attribute non-attendance could result in clarity receiving greater influence on WTP estimates relative to the other attributes, whereas WTP for other improvements in the constant baseline sample each contribute to total WTP.

Household Willingness to Pay

In the final set of comparisons we test the hypothesis that WTP is the same across baseline samples under two different illustrative regulatory scenarios. Under the first scenario the change in the environmental attributes are the same across all three baseline samples (i.e., the same delta or improvement is used to estimate benefits), but the starting points differ. In the second scenario, the policy goal is the same in each baseline, which means different improvements across baselines, as indicated in Table VIII. Expression (9) is used to calculate household WTP for both scenarios.

We use two simulation-based approaches to test for statistically significant differences in household WTP. The first, the method of convolutions, compares two empirical probability distributions by generating a third distribution for the difference between them. The probability mass below zero is a measure of the confidence level for the null hypothesis that the WTP estimates are equal (i.e. a p-value), or doubling that value for a two-tailed test. The second approach is a complete combinatorial simulation in which every pairwise combination of the simulated WTP estimates is differenced and the proportion that lies below zero provides a second estimate of the p-value for the null hypothesis that the distributions are equal. One thousand draws from a multi-variate normal distribution using the mean coefficient vector and full covariance matrix are used to generate a WTP distribution for each baseline sample. For a detailed description of how the method of convolutions is used to compare WTP distributions, and how it compares to the complete combinatorial approach and other techniques, see Poe *et al.* (1994, 2005).

Simulation results for scenario one are shown in Table VIII. Recall, in this scenario household WTP is estimated using the same change across each baseline sample. The mean household WTP for this same set of improvements is greatest in the constant baseline sample ($87 per household) and lowest in the declining baseline sample ($28). The method of convolution and the complete combinatorial approaches do not show a statistically significant difference between the constant and improving baseline results, although the improving baseline WTP is nominally lower. Household WTP is significantly lower in the declining baseline sample, however, compared to the constant baseline sample. Recall, we omit the status quo effect toward selecting a policy when estimating household WTP, and earlier comparison results suggest respondents exhibit greater attribute non-attendance and/or cognitive burdens in the non-constant baseline samples. In addition, we provide the monetized estimates for this status quo effect. These results are consistent with this this assessment. In the baseline sample respondents appear to carefully consider trade-offs across attributes, resulting in a higher household WTP for the set of attributes compared to the more complex increasing and decreasing baseline samples and their greater attribute non-attendance.

In comparison scenario two we estimate household WTP for the same policy goal in each baseline sample. Specifically, we estimate household WTP for a 10 percent improvement in each attribute from the improving baseline levels. It is necessary to specify a policy goal that is above the improving baseline to avoid extrapolating outside of the observed attribute space for all three baseline samples. All samples included choice questions with policy outcomes above the improving baseline. However, this results in some very large attribute improvements for the declining baseline sample, particularly for the clarity and low-algae lakes attributes.

The infeasibility of these changes notwithstanding, this scenario is qualitatively similar to a number of actual regulatory impact analyses. A policy goal is usually set relative to a given reference year, but uncertainty surrounding the implementation of other regulations, the models used to forecast conditions in the future, as well as changes in populations and landscapes make predicting baseline conditions difficult. To hedge against this uncertainty regulatory analyses are sometimes performed using multiple baseline scenarios (USEPA, 2015). As such, this final comparison is more useful for showing the importance of defining baseline conditions rather than how WTP functions differ across samples.

Making this comparison under our split-sample experimental design provides a useful test for external scope. Simulation results in Table IX indicate a statistically significant three-fold difference between the constant and improving baseline results, demonstrating sensitivity to scope. On average, respondents are willing to pay more for the greater improvements under the constant baseline scenario, compared to the improvements specified under the improving baseline scenario. The declining baseline WTP distribution, however, is too diffuse to make any statistical comparisons.

Finally, an important consideration when interpreting the results of this analysis is the uneven implementation of the improving baseline survey. Recall that although baseline versions of the survey were mailed evenly across all three geographic strata in the pretest, budget constraints limited administration of the improving baseline version of the main survey to only respondents in the Bay States strata (see section IV for details). To assess the robustness of our results we re-conduct the analysis using only the Bay States strata for all three baseline samples. The results are presented in Appendix B, and are discussed briefly here.

There are a few important differences in these results compared to the full sample. First, several of the regression coefficients that were previously significant are now statistically insignificant, presumably due to the smaller sample size. Second, the SQC coefficient is now of a similar magnitude across all three baselines when using the Bay States strata sample only (see Table BI). Recall, in the full sample (Table V) the SQC is more than double the magnitude in the declining baseline sample compared to the constant and improving baselines. Third, the estimates the of mean household WTP for the same change in attributes (Table BIII) suggest preferences may in fact be convex for the Bay only sample; household WTP for the same delta is smallest in the improving baseline sample and largest in the declining baseline sample.

We do find some important similarities as well. Specifically, the simulation results for the same policy goal (Table BIV) also indicate external scope sensitivity, similar to what we find when examining all geographic strata. Overall some of our baseline result comparisons are sensitive to whether the models are estimated using the full study area or focusing on just the Bay States strata. We speculate that one potential explanation for the differences in results is that respondents in the Bay States strata could be more familiar with the resource, and possibly more invested in the contingent market, than those located further from the Bay. As such, it is not surprising that these respondents exhibit preferences that in some ways appear more consistent with economic theory than the results from the larger sample. Nonetheless, our key conclusions focus on the results gleaned from the full sample across the larger study area. The larger sample size provides a more robust statistical analysis, and at the same time reveals behavioral responses across baselines that are important to highlight for future research.

VI. DISCUSSION/CONCLUSION

The baseline is a primary consideration in any benefit-cost analysis, yet has been rarely considered explicitly in stated preference research. Uncertainty about future conditions because of changing population dynamics, model uncertainty, other regulatory actions, and more can result in future conditions that are different from current ones, or what we refer to as a non-constant baseline. The presence of these factors in the Chesapeake Bay Watershed necessitated consideration of alternative baselines and afforded us the opportunity to examine how alternative baseline specifications affect WTP estimates.

We draw two primary conclusions from our results, both of which have implications for stated preference studies with non-constant baselines. First, the declining and improving baseline versions of the survey presented respondents with levels of five environmental attributes under current conditions, how conditions will change in the future under the status quo, and then changes from those future conditions dependent upon policy choices. The constant baseline version collapses the first two dimensions into a single unchanging vector of attributes over time, and in this case respondents appear to be able to better consider attribute levels in choice questions, leading to statistically significant marginal WTP estimates for each attribute. Whereas introducing changes in future baselines sometimes yields results that are inconsistent with classical economic theory. Our results suggest the non-constant baseline surveys may have required a greater cognitive burden, leading respondents to focus on one general attribute or adopt simplifying heuristics when answering the choice questions. Water clarity, notably the first (and possibly most salient) attribute in each choice question, was statistically significant across baseline samples, while most other attributes in the non-constant baseline samples were not.

In addition, respondents in all baseline samples show a strong preference for the policy options, regardless of the trade-offs among attributes, as indicated by the significance of the SQC in these samples. The SQC in the declining baseline sample is statistically larger than in the constant and improving baseline samples. It could be that respondents have an emotional reaction to declining water quality generally, and thus chose the policy option regardless of the attribute changes. Or, put differently, they substitute a simple heuristic (i.e., just do something) when faced with the more complex survey instruments.

The differences in the mixed-logit model results between the three baseline samples appear to indicate that there are either behavioral and emotional reactions or cognitive challenges associated with choice questions in the non-constant baseline surveys. If this is the case, we should proceed with caution when interpreting these results and when developing future studies in which baseline projections differ from current conditions.

Given the increased cognitive burden from non-constant baselines that appears to be present in this study, we draw several implications for future stated preference survey design. First, at the very least, practitioners should be careful in developing survey instruments where non-constant baselines are necessary, allowing for additional descriptive text and focus group testing of the provision scenario. Second, if cognitive burdens associated with non-constant baselines seem difficult to mitigate, researchers may consider reducing other dimensions of the choice task, perhaps reducing the number of attributes or alternative scenarios. Third, we recommend debriefing questions to better identify potential attribute non-attendance and cognitive difficulty, as well as respondents who may exhibit scenario rejection or other biases.

One complicating factor for interpreting the results involves the differences in endowments across respondent baseline groups. By providing different levels of future water quality, respondents across the baseline scenarios have systematically different levels of individual wealth, and therefore different reference utility levels from which we estimate MWTP. Implicit in our treatment is that preferences can be described by the same quasilinear utility function, but this may not be the case, and perhaps baseline wealth interacts with MWTP. One potentially useful future direction for research that provides alternative baseline scenarios is to explore different utility functions.

Table I: Baseline and Policy Attribute Levels

Attribute (description)	Baseline: Constant	Baseline: Declining	Baseline: Improving	Policy options
Water clarity (feet of visibility)	3	2	3.3	3; 3.5; 4.5
Striped bass population (millions of fish)	24	21	26	24; 30; 36
Blue crab population (millions of crabs)	250	235	260	250; 285; 328
Oyster population (tons)	3300	2800	4300	3300; 5500; 10,000
Lakes with low algae levels (number)	2900	2300	3100	2900; 3300; 3850
Cost (increase in annual cost of living)	$0	$0	$0	$20; $40; $60; $180; $250; $500

Table II Number of responses and response rates

	Constant	Declining	Improving	Total	Response rate[1]
Pretest	126	138	118	382	34%
Main survey	674	683	285	1,642	31%
Overall	800	821	403	2,024	

[1] Response rate is calculated based on the American Association for Public Opinion Research's Response Rate 3 calculation, which removes ineligible response plus a portion of non-responses based on an eligibility rate (AAPOR 2016).

Table III Demographic and Attitudinal Comparisons

	Constant (1)	Declining (2)	Improving (3)	p-values from t-test of means H_0: 1=3	H_0: 2=3	H_0: 1=2
Male (%)	52.4	55.4	51.6	0.22	0.64	0.37
Hispanic (%)	4.6	5.7	3.8	0.07	0.60	0.15
Black (%)	9.6	12.4	14.6	0.18	0.02	0.19
College Degree (%)	51.2	54.0	52.8	0.75	0.60	0.32
Age	56.6	55.8	55.7	0.55	0.53	0.97
Heard of the Chesapeake Bay (%)	94.5	92.8	99.2	0.00	0.00	0.23
Visited the Bay for recreation in the past 5 years (%)	36.2	32.8	65.0	0.00	0.00	0.22
Visited a Watershed Lake for recreation in the past 5 years (%)	34.3	35.0	51.5	0.00	0.00	0.78
Aware of nutrient and sediment pollution (%)	79.2	79.8	82.2	0.31	0.41	0.81

Table IV Responses to debriefing questions on scenario acceptance[1]

Debriefing Prompt	Constant	Declining	Improving
Disagreed or Strongly Disagreed with “I voted as if my household would actually face the costs shown”	5.9%	4.8% (p = 0.337)	3.9% (p = 0.144)
Disagreed or Strongly Disagreed with “I voted as if the programs would actually achieve the results shown”	6.6%	7.0% (p = 0.725)	7.9% (p = 0.378)
Agreed or Strongly Agreed with “It is important to improve the waters of the Chesapeake Bay, no matter how high the cost”	35.2%	40.0% (p = 0.060)	39.2% (p = 0.180)

[1] P-values for a two-tailed test of difference of proportions from constant baseline sample.

Table V: Mixed Logit Results by Baseline (standard deviation)

	Constant		Improving		Declining	
Variable	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
ln(clarity)	0.9263**	4.5792**	1.7289**	3.3338*	0.9716**	1.9822
	(0.4697)	(0.7773)	(0.7143)	(1.7574)	(0.4151)	(2.8176)
ln(bass)	1.2412**	2.7088**	0.1301	3.2140	0.3453**	-2.7137
	(0.4258)	(1.1016)	(0.7693)	(2.0780)	(0.4319)	(1.7940)
ln(crab)	2.2716**	-0.5751	1.5509*	3.4340	-0.1007	1.8597
	(0.6147)	(2.1489)	(0.8777)	(2.9679)	(0.6306)	(3.6904)
ln(oyster)	0.3708**	0.4172	0.2102	0.1120	0.1140	-0.8411**
	(0.1490)	(0.5909)	(0.2510)	(0.4740)	(0.1488)	(0.3981)
ln(lakes)	3.5394**	3.6437**	1.0713	1.1138	0.3075	-3.5644
	(0.6390)	(1.3548)	(1.6386)	(2.8742)	(0.5383)	(2.4625)
Cost	-0.0092**		-0.0082**		-0.0077**	
	(0.0008)		(0.0082)		(0.0007)	
SCQ	-1.9958**	4.5278**	-1.8910**	4.3531**	-3.9508**	4.3164**
	(0.3448)	(0.4372)	(0.4989)	(0.5941)	(0.5804)	(0.6307)
Observations	5,103		2,493		5,256	
Respondents	605		287		614	

* significant at the 0.1 level ** significant at the 0.05 level

Table VI: Marginal WTP by Baseline (standard deviation); $2016

Scenario	Bay Water Clarity	Striped Bass Population	Blue Crab Population	Oyster Abundance	Low Algae Lakes
Constant	2.81*	5.65**	0.99**	0.01**	0.13**
	(1.40)	(1.94)	(0.26)	(0.00)	(0.02)
Improving	5.83**	0.66	0.75*	0.01	0.04
	(2.37)	(3.89)	(0.42)	(0.01)	(0.07)
Declining	5.25**	2.13	-0.05	0.01	0.02
	(2.06)	(2.64)	(0.33)	(0.01)	(0.03)

* significant at the 0.1 level ** significant at the 0.05 level

Table VII: Attribute Improvements Under Each Comparison Scenario

Baseline	Bay Water Clarity (inches)	Striped Bass Population (million fish)	Blue Crab Population (million crab)	Oyster Abundance (tons)	Low Algae Lakes
			Same Improvement		
All Three Baselines	3.6	2.4	25	330	290
			Same Policy Goal		
Constant	7.56	4.6	36	1,430	510
Improving	3.96	2.6	26	430	310
Declining	19.56	7.6	61	1,930	1,110

Table VIII: Comparison Scenario 1: Same Improvements for Each Attribute by Baseline

			Significance of difference from Constant Baseline WTP	
	Mean Household WTP	95% Confidence Interval	Method of Convolution	Complete combinatorial
Constant	$87**	[62 - 115]		
Improving	$51*	[2 - 100]	0.209	0.211
Declining	$28	[-8 - 66]	0.010	0.010

* significant at the 0.10 level ** significant at the 0.05 level

Table IX: Comparison Scenario 2: Same Policy Goals for Each Attribute by Baseline

			Significance of difference from Constant Baseline WTP	
	Mean Household WTP	95% Confidence Interval	Method of Convolution	Complete combinatorial
Constant	$154**	[107 - 202]		
Improving	$56*	[-2 - 113]	0.007	0.008
Declining	$109	[-32 - 228]	0.497	0.499

* significant at the 0.1 level ** significant at the 0.05 level

References

Abt Associates, Inc. 2016. Annotated Bibliography of Surface Water Valuation Studies. January 13, 2016.

Alpizar, Francisco, Carlsson, Fredrik, Martinsson, Peter. 2001. Using Choice Experiments for Non-Market Valuation. *Econ. Issues* 8(1): 83-110.

American Association for Public Opinion Research (AAPOR). 2016. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf.

Banzhaf, H. Spencer, Dallas Burtraw, David Evans and Alan Krupnick. 2006. Valuation of Natural Resource Improvements in the Adirondacks. *Land Economics,* 82(3): 445-464.

Bennett, Jeff, Adamowicz, Vic. 2001. Some Fundamentals of Environmental Choice Modelling. In Bennett, Jeff, Blamey, Russell. (Eds.) *The Choice Modelling Approach to Environmental Valuation.* Massachusetts: Edward Elgar Publishing Limited.

Boxall, Peter, Adamowicz, W.L. (Vic), Moon, Amanda. 2009. Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement. *Aust. J. of Agric. and Resour. Econ.*, *53*(4): 503-519.

Cameron, Trudy Ann, DeShazo, J.R., Johnson, Erica H., 2011. Scenario adjustment in stated preference research. *J. of Choice Modelling*, *4*(1): 9-43.

Carson, Richard T., Czajkowski, Mikolaj. 2014. The Discrete Choice Experiment Approach to Environmental Contingent Valuation. In. Hess, Stephane, Daly, Andrew. (Eds.) *Handbook of Choice Modeling.* Northampton: Edward Elgar Publishing.

Cummings, Ronald G., Ganderton, Philip T., McGuckin, Thomas. 1994. Substitution Effects in

CVM Values. *Am. J. of Agric. Econ.* 76: 205-214.

Daly, Andrew, Hess, Stephane, Train, Kenneth. 2012. Assuring Finite Moments for Willingness to Pay in Random Coefficients Models. *Transportation* 39(1): 267-297.

Dillman, Donald A. 2008. *Mail and Internet Surveys: The Tailored Design Method*. New York: John Wiley and Sons.

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

Johnston, Robert J., Boyle, Kevin J., Adamowica, Wiktor (Vic), Bennett, Jeff, Brouwer, Roy, Cameron, Trudy Ann, Hanemann, W. Michael, Hanley, Nick, Ryan, Mandy, Scarpa, Ricardo, Tourangeau, Roger, Vossler, Christian A. 2017. Contemporary Guidance for Stated Preference Studies. *Journal of the Association of Environmental and Resource Economists* 4(2): 319-405.

Johnston, Robert J., Swallow, Stephen K., Tyrrell, Timothy J., Bauer, Dana Marie. 2003. Rural Amenity Values and Length of Residency. *Am. J. of Agric. Econ.* 85(4): 1009-1024.

Lew, Daniel K., Layton, David F., Rowe, Robert D. 2010. Valuing Enhancements to Endangered Species Protection Under Alternative Baseline Futures: The Case of the Steller Sea Lion. *Marine Resour. Econ.* 25*: 133-154.*

Massey, David M., Moore, Chris, Newbold, Stephen C., Ihde, Tom, Townsend, Howard. 2017. A Commercial Fishing and Outdoor Recreation Benefits of Water Quality Improvements in the Chesapeake Bay. *US EPA National Center for Environmental Economics Working Paper* 17-02*2.*

McFadden, Daniel, Train, Kenneth. 2000. Mixed MNL Models for Discrete Response. *J. of Applied Econometrics* 15: 447-470.

Moore, Richard B., Johnston, Craig M., Smith, Richard A., Milstead, Bryan. 2011. Source and Delivery of Nutrients to Receiving Waters in the Northeastern and Mid-Atlantic Regions of the United States. *J. of the Am. Water Resour. Assoc. 47*(5): 965-990.

Moore, Chris, Guignet, Denny, Maguire, Kelly, Dockins, Chris, Simon, Nathalie. 2015. A Stated Preference Study of the Chesapeake Bay and Watershed Lakes. *US EPA National Center for Environmental Economics Working Paper* 15-06. https://www.epa.gov/sites/production/files/2016-03/documents/2015-06.pdf

Moore, Chris, Guignet, Dennis; Kelly B. Maguire; Chris Dockins; and Nathalie B. Simon. 2018. Valuing Ecological Improvements in the Chesapeake Bay and the Importance of Ancillary Benefits. *J. of Benefit Cost Analysis*. forthcoming.

Poe, Gregory, Giraud, Kelly L., Loomis, John. 2005. Computational Methods for Measuring the Difference of Empirical Distributions. *Am. J. of Agric. Econ.* 87(2): 353-365.

Poe, Gregory, Welsh, Michael P., Severance-Lossin, Eric K. 1994. Measuring the Difference (X-Y) in Simulated Distributions: Application of the Convolutions Approach. *Am. J. of Agric. Econ.* 76(4): 904-915.

Soto Montes deOca, Gloria, Bateman, Ian J. 2006. Scope Sensitivity in Households' Willingness to Pay for Maintained and Improved Water Supplies in a Developing World Urban Area: Investigating the Influence of Baseline Supply Quality and Income Distribution Upon Stated Preferences in Mexico City. *Water Resour. Research* 42(7): 1-15.

Swait, Joffre, Louviere, Jordan. 1993. The Role of the Scale Parameter in the Estimation and

Comparison of Multinomial Logit Models. *J. of Marketing Research* 30(3): 305-314.

U.S. Environmental Protection Agency (USEPA). 2010a. *Guidelines for Preparing Economic Analysis*. EPA-240-R-10-001, December.

USEPA. 2010b. *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.

USEPA. 2015. Regulatory Impact Analysis for the Effluent Limitations Guidelines and Standards for the Steam Electric Power Generating Point Source Category, Office of Water. EPA-821-R-15-004.

Appendix A. Unscreened Sample

Table AI: Mixed Logit Results using Unscreened Sample by Baseline (standard deviation)

	Constant		Improving		Declining	
Variable	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
ln(clarity)	0.7803	5.7056***	1.8447**	5.2568***	0.5553	2.1904
	(0.4747)	(0.0972)	(0.7349)	(1.2805)	(0.4652)	(1.9361)
ln(bass)	1.1440***	3.0038**	0.3220	3.0456	0.2433	3.3765***
	(0.4189)	(1.4460)	(0.7843)	(3.2972)	(0.3861)	(1.0383)
ln(crab)	2.1566***	0.1125	1.6863**	0.8479	0.0897	0.1084
	(0.5869)	(1.4267)	(0.8080)	(2.8774)	(0.5810)	(1.7646)
ln(oyster)	0.2905*	1.0100	0.2219	0.6429	0.0695	-0.7592
	(0.1498)	(0.6384)	(0.2409)	(0.4986)	(0.1309)	(0.7892)
ln(lakes)	3.3971***	3.0877**	-0.0195	8.3409***	0.3384	3.1174
	(0.6318)	(2.1133)	(1.6633)	(2.9110)	(0.4902)	(2.0111)
Cost	-0.0079***		-0.0089***		-0.0066***	
	(0.0010)		(0.0009)		(0.0007)	
SCQ	-0.9498***	5.9919***	-0.9459*	5.8384***	-3.4651***	5.8594***
	(0.3228)	(0.5130)	(0.5063)	(0.6027)	(0.6225)	(0.5539)
Observations	6,795		3,447		6,807	
Respondents	796		395		792	

* significant at the 0.1 level   ** significant at the 0.05 level

Table AII: Marginal WTP Estimates using Unscreened Sample by Baseline (standard deviation).

Scenario	Bay Water Clarity	Striped Bass Population	Blue Crab Population	Oyster Abundance	Low Algae Lakes
Constant	2.43*	5.34***	0.97***	0.01*	0.13***
	(1.47)	(1.98)	(0.25)	(0.01)	(0.02)
Improving	5.90**	1.57	0.82**	0.01	-0.01
	(2.34)	(3.78)	(0.39)	(0.01)	(0.07)
Declining	3.51	1.76	0.06	0.00	0.02
	(2.76)	(2.76)	(0.39)	(0.01)	(0.03)

* significant at the 0.1 level   ** significant at the 0.05 level

Appendix B: Bay States Only Sample

Table BI: Bay States Only Mixed Logit Results by Baseline (standard deviation)

	Constant		Improving		Declining	
Variable	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
ln(clarity)	1.1543	3.3663**	2.144**	3.8165	1.5943**	-4.2529**
	(0.7719)	(1.3256)	(0.8860)	(3.4264)	(0.7014)	(1.9005)
ln(bass)	0.9621	-3.0231	-0.554	3.8790	1.1692	3.4944**
	(0.7211)	(1.9775)	(0.8914)	(2.5025)	(0.7521)	(1.3763)
ln(crab)	0.9402	-0.6811	1.568	2.6039*	1.2632	6.1133
	(1.0438)	(3.1163)	(1.0354)	(1.4652)	(1.4261)	(6.8957)
ln(oyster)	0.4467**	0.1790	0.319	-0.1069	0.4935	1.0743
	(0.2504)	(0.4656)	(0.2850)	(0.2921)	(0.3426)	(1.0954)
ln(lakes)	3.1634**	-6.4128**	-0.774	-0.8124	1.1285	1.1943
	(1.0346)	(2.9744)	(1.9863)	(3.2128)	(0.9258)	(0.9928)
Cost	-0.0087**		-0.009**		-0.0083**	
	(0.0013)		(0.0011)		(0.0014)	
SCQ	-2.3632**	4.3868	-2.668**	4.9821**	-2.6619**	3.4181**
	(0.5423)	(0.8680)	(0.6994)	(0.8547)	(0.9707)	(1.0793)
Observations	1,879		1,938		1,944	
Respondents	224		222		227	

Table BII: Bay States only Marginal WTP by Baseline (standard deviation); $2016

Scenario	Bay Water Clarity	Striped Bass Population	Blue Crab Population	Oyster Abundance	Low Algae Lakes
Constant	3.67	4.59	0.43	0.02*	0.12**
	(2.36)	(3.39)	(0.47)	(0.01)	(0.04)
Improving	6.26**	-2.46	0.70	0.01	-0.03
	(2.56)	(3.99)	(0.45)	(0.01)	(0.07)
Declining	7.99**	6.70	0.68	0.02*	0.06
	(3.45)	(3.99)	(0.72)	(0.01)	(0.04)

Table BIII. Comparison Scenario 1: Same Improvements for Each Attribute by Baseline for the Bay States Only

			Significance of difference from Constant Baseline WTP	
	Mean Household WTP	95% Confidence Interval	Method of Convolution	Complete combinatorial
Constant	$72	[21 - 122]		
Improving	$25	[-30 - 73]	0.220	0.222
Declining	$82	[7-158]	0.861	0.863

* significant at the 0.1 level ** significant at the 0.05 level

Table BIV. Comparison Scenario 2: Same Policy Goals for Each Attribute by Baseline for the Bay States Only

			Significance of difference from Constant Baseline WTP	
	Mean Household WTP	95% Confidence Interval	Method of Convolution	Complete combinatorial
Constant	$130	[32 - 219]		
Improving	$33	[-39 - 102]	0.082	0.084
Declining	$301	[68 - 531]	0.183	0.184

* significant at the 0.1 level ** significant at the 0.05 level