PART B OF THE SUPPORTING STATEMENT

1.   Survey Objectives, Key Variables, and Other Preliminaries

1(a)   Survey Objectives

The overall goal of this survey is to examine the total value of benefits (including non-use values) for improvements in water quality in the Chesapeake Bay and its Watershed. Water quality improvements are expected to follow nitrogen, phosphorous, and sediment load reductions set forth in recent Chesapeake Bay Total Maximum Daily Load (TMDL) requirements. EPA has designed the survey to provide data to support the following specific objectives:

- To estimate the total values, including non-use values that individuals place on improving water quality in the Chesapeake Bay and lakes in the Watershed.
- To understand how individuals value improvements in the Chesapeake Bay and lakes in the Watershed, including: water clarity; populations of striped bass, blue crab, and oysters; and lake conditions.
- To understand how the above values depend on the future baseline level of water quality in the Chesapeake Bay and its Watershed.
- To understand how values vary with respect to individuals' attitudes, awareness, and demographic characteristics.

Understanding total public values for water quality improvements is necessary to determine the full range of benefits associated with reductions in nutrient (nitrogen and phosphorous) and sediment loads to the Chesapeake Bay. While direct use values can be estimated using a variety of methods, non-use values can only be assessed via stated preference survey methods. Because non-use values may be substantial, failure to recognize such values may lead to improper inferences regarding policy benefits (Freeman 2003).

1(b)   Key Variables

The key questions in the survey ask respondents whether or not they would vote for policies that would result in improvements in water quality in the Chesapeake Bay and lakes in the Watershed in exchange for an increase in their cost of living. The choice experiment framework allows respondents to view pairs of multi-attribute policies associated with total maximum daily loads to the Chesapeake Bay. Respondents are asked to choose one of three options. Two of these options correspond to "additional programs" that yield improvements in some or all of the environmental attributes specified, and the third option is the status quo (i.e., maintain current programs with no additional household costs). The survey design follows well-established choice experiment methodology and format (Adamowicz et al. 1998; Louviere et al. 2000; Bennett and Blamey 2001; Bateman et al. 2002).

The survey focuses on environmental and ecological "endpoints." In other words, it asks respondents about changes in attributes that directly enter into their household production and utility function. Specifically, the survey presents changes in the following attributes: (a) water clarity, (b) striped bass population, (c) blue crab population, (d) oyster abundance, (e) conditions of freshwater lakes in the Watershed, and (e) the cost of living.   As discussed by Boyd and Krupnick (2009), these endpoints are aspects of the environment that people experience, make choices about, and have a tangible meaning.

The study design includes three treatment levels in which environmental conditions without additional action are either declining (i.e., "declining baseline"), unchanged (i.e., "constant baseline"), or improving ("improving baseline"), relative to the conditions today. As discussed in Part A of this ICR, section 4(b) (i), the three survey versions are included because there is uncertainty regarding how population growth and changes in land use patterns will affect water quality in the future. There is also the question of what practices that are not in place now would be in place in 2025 if the TMDLs were not implemented. . Given that there will be some uncertainty regarding the specifics of the "actual" baselines and improvements, the resulting valuation estimates will allow flexibility in estimating WTP for a range of different circumstances. It also increases the potential that the results of the study can be used in other studies as part of a benefit transfer exercise, consistent with EPA's *Guidelines for Preparing Economic Analyses* (US EPA 2010). The three versions of the survey are included in this document as Attachments 1 through 3.

While EPA intends to administer all three baseline versions of the survey in all three geographic strata, budget considerations may cause that plan to be scaled back so that not all versions are administered in all strata.

The analysis of choice questions will use data on how the respondent votes, the amount of the cost of living increase, and the degree of improvement in the environmental attributes, to estimate values for changes in those attributes. Variables for socio-economic characteristics and attitudes will also be included in the analysis.

1(c)   Statistical Approach

A statistical survey approach in which a randomly drawn sample of households is asked to complete the survey is appropriate for estimating the use and non-use values associated with the Chesapeake Bay and its Watershed. A census approach is impractical because of the extraordinary cost of contacting all households. The relevant population includes households not only residing near the Chesapeake Bay, but more distant states on the east coast. Specifically, the sample population includes households in the Bay states (MD, VA, DC), Watershed states (DE, NY, PA, WV) and other East Coast states (CT, FL, GA, MA, ME, NC, NH, NJ, RI, SC, VT). An alternative approach, where individuals self-select into the sample, is not sufficiently rigorous to provide a useful estimate of the total value of water quality and habitat improvements. Therefore the statistical survey is the most reasonable approach to estimate the total value of the Chesapeake Bay TMDLs.

Much of the work in developing the survey instrument was conducted by the EPA, and EPA will also directly conduct much of the analysis of the survey results. EPA has retained Abt Associates Inc. (55 Wheeler Street, Cambridge, MA 02138) under EPA contract EP-W-11-003 to assist in the questionnaire design, sampling design, administration of the survey, and analysis of the survey results.

1(d)   Feasibility

Following standard practice in the stated preference literature (Adamowicz et al. 1998; Batemen, et al, 2002; Bennett and Blamey 2001; Johnston et al. 1995; Louviere et al. 2000), EPA conducted a series of 10 focus groups and an initial set of 26 cognitive interviews (OMB control # 2090-0028).

Based on findings from these activities, EPA made various improvements to the survey instrument to reduce the potential for respondent bias, reduce respondent cognitive burden, and increase respondent comprehension of the survey materials. In addition, EPA solicited peer review of the survey instruments by three specialists in academia, as well as input from other experts (see section 3c in Part A). Recommendations and comments received as part of that process have been incorporated into the design of the survey instrument and the revised survey was subsequently tested in an additional 46 cognitive interviews.

Because of the steps taken during the survey development process, EPA does not anticipate that respondents will have difficulty interpreting or responding to any of the survey questions. Furthermore, since the survey will be administered as a mail survey, it will be easily accessible to all respondents. EPA therefore believes that respondents will not face any obstacles in completing the survey, and that the survey will produce useful results. EPA has dedicated sufficient staff time and resources to the design and implementation of this survey, including funding for contractor assistance under EPA contract No. EP-W-11-003. Given the timetable outlined in Section A 5(d) of this document, the survey results should be available for timely use in the final benefits analysis for the Chesapeake Bay TMDLs.

2. Survey Design

2(a) Target Population and Coverage

To assess both use- and non-use values for improvements in Chesapeake Bay water quality, the target population is individuals who are 18 years of age or older and reside in the District of Columbia or one of 17 east coast U.S. states: Maryland, Virginia, Delaware, New Jersey, New York, Pennsylvania, West Virginia, Vermont, New Hampshire, Massachusetts, Connecticut, Rhode Island, Maine, North Carolina, South Carolina, Georgia, or Florida. These were chosen based on their immediate proximity to the Bay (Maryland, Virginia, District of Columbia) and/or lakes, streams and rivers in its Watershed (Delaware, New York, Pennsylvania, West Virginia). 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 remaining states (i.e., Vermont, New Hampshire, New Jersey, Massachusetts, Connecticut, Rhode Island, Maine, North Carolina, South Carolina, Georgia, and Florida) 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 and isolate non-use values.

2(b)   Sampling Design

(i)   Sampling Frame

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

For discussion of techniques that EPA will use to minimize non-response and other non-sampling errors in the survey sample, refer to Section 2(b)(II), below.

(ii)   Sample Sizes

The target responding sample size for the survey is 2,592 completed household surveys. 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 (see section 2(c)) achieved by the analysis will be more than adequate to meet the analytic needs of the benefits analysis for the

_____

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

Chesapeake Bay TMDLs. For further discussion of the level of precision required by this analysis, see Section 2(c)(i) below.

The sample design features three geographic strata based upon proximity to the Chesapeake Bay and its Watershed: Bay States, Watershed States, and East Coast States. Relative to the geographic distribution of households across the East Coast, EPA plans to over-sample households in states adjacent to Chesapeake Bay and within the Chesapeake Bay Watershed. EPA believes this approach is appropriate because households in these areas will incur the costs of Chesapeake Bay water quality improvements, and therefore this group is most likely to receive use-value benefits from the improvements. Within each survey region, the household sample will be allocated in proportion to the geographic distribution of households within states in the three regions.   The target number of responding households in each region and state is given below (Table B1). For discussion of the required sample size by state, please refer to Error: Reference source not found.

Table B1. Total Households and Expected Number of Completed Surveys for Each Study Region.		
Sampling Stratum and State	**Total Number of Households**	**Expected Number of Completed Surveys**
Bay States Stratum, total	5,479,176	864
District of Columbia	266,707	42
Maryland	2,156,411	340
Virginia	3,056,058	482
Watershed States Stratum, total	13,442,787	864
Delaware	342,297	22
New York	7,317,755	470
Pennsylvania	5,018,904	323
West Virginia	763,831	49
Other East Coast States Stratum, total	25,431,478	864
Connecticut	1,371,087	47
Florida	7,420,802	252
Georgia	3,585,584	122
Maine	557,219	19
Massachusetts	2,547,075	87
New Hampshire	518,973	17
New Jersey	3,214,360	109
North Carolina	3,745,155	127
Rhode Island	413,600	14

South Carolina	1,801,181	61
Vermont	256,442	9
Total	44,353,441	2,592

Source: U.S. Census Bureau (2012). 2010 Census Summary File 1. Retrieved May 31, 2012 from http://factfinder2.census.gov/.

(iii)   Stratification Variables

The population of households in the eastern United States is stratified by the geographic boundaries of three study regions in Table B1: states adjacent to Chesapeake Bay ("Bay States"), states which contain the Chesapeake Bay Watershed ("Watershed States"), and additional East Coast states ("East Coast States"). Bay States include MD, VA, and DC; Watershed States include DE, NY, PA, and WV; and East Coast States include VT, NH, NJ, MA, CT, RI, ME, NC, SC, GA, and FL.

The sample will be allocated in equal proportions of 33% for each stratum, thus leading to the highest sampling rate in the Bay State stratum, and the lowest sampling rate in the Other East Coast States stratum. The expected number of completed interviews in each stratum is presented in Table B2. This allocation is designed to minimize the variance of the main effects under an experimental design model as introduced in Section 4(a) of Part A. As a result of stratification, the analysis will produce estimates of the geographic distribution of values for the Chesapeake Bay water quality improvements with greater precision.

(iv)   Sampling Method

Using the stratification design discussed above, sample households will be randomly selected from the U.S. Postal Service DSF database. Assuming 92% of the sampled addresses are eligible and 30% of eligible households will return a completed mail survey, 9,391 households will be sampled from the DSF.[2]

---

[2] A 92% eligibility rate is based on Link, et al. (2008). A 30% response rate is based on Helm (2012,), Mansfield, et al. (2012), and Johnston, et al. (2012).

For obtaining population-based estimates of various parameters, each responding household will be assigned a sampling weight. The weights will be used to produce estimates that:

- are generalizable to the population from which the sample was selected;
- account for differential probabilities of selection across the sampling strata;
- match the population distributions of selected demographic variables within strata; and
- allow for adjustments to reduce potential nonresponse bias.

These weights combine:

- a base sampling weight which is the inverse of the probability of selection of the household;
- a within-stratum adjustment for differential non-response across strata; and
- a nonresponse weight.

Post-stratification adjustments may be made to match the sample to known population values (e.g., from Census data).

There are various models that can be used for nonresponse weighting. For example, nonresponse weights can be constructed based on estimated response propensities or on weighting class adjustments. Response propensities are designed to treat nonresponse as a stochastic process in which there are shared causes of the likelihood of nonresponse and the value of the survey variable. The weighting class approach assumes that within a weighting class (typically demographically-defined), non-respondents and respondents have the same or very similar distributions on the survey variables. If this model assumption holds, then applying weights to the respondents reduces bias in the estimator that is due to nonresponse. Several factors, including the difference between the sample and population distributions of demographic characteristics, and the plan for how to use weights in the regression models will determine which approach is most efficient for both estimating population parameters and for the stated-preference modeling.

To estimate total value for the quantified environmental benefits of the Chesapeake Bay TMDLs, data will be analyzed statistically using a standard random utility model framework.

(v)   Multi-Stage Sampling

Multi-stage sampling will not be necessary for this survey.

2(c)   Precision Requirements

(i)   Precision Targets

Table B2 presents expected sample sizes for each geographic stratum. The maximum acceptable sampling error for predicting response probabilities (i.e., the likelihood of choosing a given alternative) in the present case is ±10%, assuming a true response probability of 50% associated with a utility indifference point. Given the survey population size, this level of precision requires a minimum sample size of approximately 96 observations. The number of observations (i.e., completed surveys) required to obtain large sample properties for the choice experiment design provide more than sufficient observations to obtain this required precision for population parameters. Across all regions, a sample of 2,592 households (completed surveys) will provide estimates of population percentages with a level of precision ranging from 1.1% at the 50% incidence level to 0.7% at the 10% incidence level (Table B2).

Table B2. Sample size and accuracy projections.

Geographic division	Population size	Expected sample size (completed surveys)	Expected weights	Standard error, 50% incidence	Standard error, 10% incidence
Bay States	5,479,176	864	6,341	0.017	0.010
Watershed	13,442,787	864	15,559	0.017	0.010
Other East Coast	25,431,478	864	29,434	0.017	0.010
Overall	44,353,441	2,592	17,081	0.011	0.007

Source for household population size: U.S. Census Bureau (2012). 2010 Census Summary File 1. Retrieved May 31, 2012 from http://factfinder2.census.gov/. The margin of error is 1.96 times the standard error.

(ii)   Power analysis

Power analysis in this section is performed for a one-sample t-test of proportions for study as a whole. The accuracy of the WTP estimates, and hence the power to detect differences in WTP, depends on the true values of the parameters of the logistic model used in WTP estimation, and hence can only be conducted post-hoc after the parameter estimates are obtained. Given the stratified nature of the survey, the variance of a z-test statistic comparing the null incidence p_0 with the alternative incidence p_1 is

$$V(p,n)=\frac{\sum_{h=1}^{3} W_h^2 p(1-p)(1+CV^2)^{\square}}{n}$$

where CV is the coefficient of variation of post-stratified weights within a stratum $1+CV^2$ is the design effect due to variable weights, or 'DEFF'), $n=864$ is the target number of completed surveys for a given stratum, and W_h is the proportion of the population in stratum $h=1$ (Bay States), 2 (Watershed), 3 (Other East Coast):

$$W_1=\frac{5,479,176}{44,353,441}=12.3\%, W_2=\frac{13,442,787}{44,353,441}=30.3\%, W_3=\frac{25,431,478}{44,353,441}=57.3\%$$

Thus, $\sum_h W_h^2=0.4359\sum_h W_h^2=0.4359$. For a given power level (e.g.. 80%), the effect size that can be determined by solving

$$p_1-p_0=z_{1-\alpha/2} V^{1/2}(p_0,n)+z_{1-\beta} V^{1/2}(p_1,n)$$

for p_1. This is an extension of the standard power analysis for stratified samples.

Table B3 lists effect sizes using the most typical values for significance level ($\alpha=5\%$) and power ($\beta=80\%$), and for various scenarios concerning variability of weights within strata (which will be caused by differential non-response). Across these scenarios, the margin of error on an estimate of population proportions[3] ranges from 1.72 to 2.84 percentage points at the 95% confidence level.

Table B3. Power analysis.

Effect size detectable with power 80% by a test of size 5%	$p_0 = 50\%$	$p_0 = 10\%$
CV = 0.32, within-strata DEFF due to weights = 1.1	2.61%	1.58%
CV = 0.55, within-strata DEFF due to weights = 1.3	2.84%	1.72%

(iii) Non-Sampling Errors

---

[3] i.e., Whether it is true that the target population contains 50% women ($p_0=50\%$), or 10% women ($p_0=10\%$).

Several non-sampling errors may be encountered in stated preference surveys. First, protest responses can occur when individuals reject the survey format or question design, even though they may value the resources being considered (Mitchell and Carson 1989).   To help identify protest responses EPA has included several survey debriefing questions, e.g., whether the respondent would oppose any government program that imposes more regulation and spending (see section 4(b) (i) in Part A of this ICR for details). The use of such methods to identify protest responses is well-established in the literature (Bateman et al. 2002). Moreover, researchers (e.g., Bateman et al. 2002) suggest that a choice experiment format, such as that proposed here, may ameliorate such responses (as opposed to say, a contingent valuation format).

Non-response bias is another type of non-sampling error that can potentially occur in stated preference surveys. Non-response bias can occur when households choose not to participate in a survey (i.e., not return the mail survey, in this case) or do not answer all relevant questions on the survey instrument. EPA has designed the survey instrument to maximize the response rate. EPA will also follow Dillman's (2008) mail survey approach (see subsection 4(b) for details). If necessary, EPA will use appropriate weighting or other statistical adjustments to correct for any bias due to non-response.

To determine whether there is any evidence of significant non-response bias in the completed sample, EPA will conduct a non-response follow-up study. This will enable EPA to identify potential differences between respondents to the mail survey and those who received a questionnaire but did not return it.

Non-response Follow-up Survey

In order to ascertain if and how respondents and non-respondents differ EPA will administer a short non-response follow-up survey to a random sample of households that receive the main survey but do not complete and return it.   The short questionnaire will ask a few awareness, attitudinal and demographic questions that can be used to statistically examine differences, if any, between respondents and non-respondents. It will take respondents about 5 minutes to complete the non-response follow-up survey. The short questionnaire will be implemented using priority mailing contact. The samples for the non-response follow up will be allocated proportionately to the number of the original mailings in the geographic division (state). The mailing will include $2 in cash as an unconditional incentive for completion of the short questionnaire to encourage response.

EPA expects a 20% response rate. Based on the response rate observed in the nonresponse survey pretest study, EPA will send (via priority mail) the nonresponse survey to enough households in order to obtain the expected target sample of 720 completed nonresponse questionnaires. This sample size will enable EPA to reject the hypothesis of no difference in population percentages between respondents and non-respondents with 80% power when there is a difference of 10.1% according to a two-sided statistical test at the base incidence of 50%, or a difference of 6.9% at the base incidence of 10%. Table B4 illustrates the target sample size of the nonresponse survey across survey regions.

Table B4: Expected Number of Completed Non-response Follow-up (NRFU) Surveys	
Region	**Number of Households Expected to Return Priority Mail NRFU Survey**
Bay	240
Watershed	240
Other East Coast	240
Total	720

2(d)   Questionnaire Design

The information requested by the survey is discussed in Section 4(b)(I) of Part A of the supporting statement. The full texts of the three draft questionnaires are provided in Error: Reference source not found, 2, and 3.

Several categories of questions are included in the survey. The reasons for including each of these categories are discussed below:

- **Familiarity with the Chesapeake Bay Watershed and Watershed Issues** (pgs. 1 to 2)**.** Responses to these questions provide information on whether respondents visited or viewed Chesapeake Bay and lakes in the Watershed. These questions will allow EPA to identify respondents who are users of the resource or non-users. Identifying non-users is important for

estimating non-user WTP. Additionally, respondents who have increased contact with the aquatic resources are expected to have higher values for improvements in environmental conditions (Johnston et al. 2005). The questions in this section will identify whether respondents' recreational experiences have included the Chesapeake Bay or lakes in the Watershed or both. This section also provides respondents with some background information on nutrient and sediment pollution issues in the Chesapeake Bay Watershed. Responses to question 4 will be used to assess respondents' a priori knowledge about these issues, and potentially test for knowledge effects on respondent answers to the program choice questions. Respondents who are more familiar with these pollutants may have different values, although the direction of such effects is unknown a priori.

- **Current and Future Conditions in the Chesapeake Bay and Watershed Lakes** (pgs. 3 to 4)**.** In this section respondents are presented with the environmental attributes that are applied in the later choice questions: bay water clarity; striped bass, blue crab, and oyster populations; and the number of Watershed lakes with relatively low algae levels. Respondents are given the current levels of these attributes, along with the levels predicted for 2025 under one of three baseline assumptions: declining, constant, or improving. Responses to question 5 gauge respondents' expectations regarding future environmental conditions in the Chesapeake Bay and Watershed lakes (absent additional pollution control actions).

- **Additional Pollution Programs for the Chesapeake Bay Watershed** (pg 5 to 6)**.** This section introduces the concept that additional pollution reduction programs are being considered. Question 6 inquires about respondents' knowledge of whether they currently pay any environmentally related taxes or fees. Responses to this question can be used to examine whether respondents already paying for environmental improvements are less willing to pay additional amounts for further environmental improvements. This section also provides text introducing the payment vehicle for the later choice questions. Additional text is included to further emphasize the consequentiality of one's choices. In order to minimize consideration of omitted variables, this section also provides text encouraging respondents to only consider the environmental outcomes presented in the choice questions, thus reducing the potential for omitted variable bias.

- **Deciding Future Actions (**pgs 7 to 8)**.** This section provides instructions and an example of how one responds to the choice questions. This section also includes "cheap talk" text emphasizing

that even though this exercise is hypothetical, respondents should consider their choices as if they were real.

- **Voting for Programs to Improve the Condition of Chesapeake Bay and lakes in the Watershed** (pgs. 9 to 11)**.** The questions in this section are the key component of the survey. Respondents' choices among alternatives with specific environmental quality improvements and household cost increases are the main data that allow estimation of willingness to pay. The questions are presented in a choice experiment, where respondents choose their preferred option: option A (status quo), option B, and option C. This elicitation format has been successfully used by a number of previous valuation studies (e.g., Adamowicz et al. 1998; Bateman et al. 2002; Bennett and Blamey 2001; Louviere et al. 2000; Johnston et al. 2002a, 2005, 2012; Opaluch et al. 1993).

- **Debriefing Questions** (pg. 12)**.** These questions ask respondents about their motivations as to why they chose certain choice options over others, and whether they accepted the hypothetical scenario when making their choices. These questions will help to identify respondents who incorrectly interpreted the choice questions or did not believe the choice scenarios to be credible. In other words, the responses to these questions will be used to identify potentially invalid responses, such as: protest response (e.g., protest any government program), scenario rejection, omitted variable considerations (e.g., availability and/or price of seafood, economy and employment, recreation in lakes outside the Watershed, etc.), and symbolic (warm glow) responses. Some questions in this section are also included in the non-response follow-up survey, which allows for an examination of whether there are any statistical differences between respondents and non-respondents. The responses to some of the questions will also be used to identify use versus nonuse motivations behind respondents' choices, including: altruism, option value, bequest, etc.

- **Recreational Experience and Time Preferences** (pg 13). Responses to these questions elicit recreational experience data to test if certain respondent characteristics influence responses to the choice questions. These questions will also allow EPA to identify resource non-users for purposes of estimating nonuser WTP, and therefore to gauge the relative importance of non-use values to overall benefits. The questions will also identify frequency of use and whether the respondent's recreational experiences have included lakes in the Watershed, the Chesapeake Bay

itself, or both. Question 15 is included to elicit information on respondents' internal discount rate.

- **Demographics** (pg. 14)**.** Responses to these questions will be used to estimate the influence of demographic variables on respondents' voting choices, and ultimately, their values to improve environmental quality in the Chesapeake Bay and lakes in the Watershed. By including this page in both the survey and non-response follow-up survey, statistical comparisons of household characteristics can be across the samples of responding and nonresponding households. These data can also be compared to household characteristics from the sample frame population, which are available from the 2010 Census.

3.   Pretests

EPA conducted extensive testing of the survey instrument during a set of 10 focus groups and 72 cognitive interviews (OMB Control Number 2090-0028). Individuals in these focus groups participated in discussions about the Chesapeake Bay and its Watershed. They also completed draft survey questionnaires and provided comments and feedback about the survey format and content, their interpretations of the questions, and other issues relevant to stated preference estimation. Individual cognitive interviews with survey respondents were conducted using think-aloud or verbal protocol analyses (see Schkade and Payne 1994 for a discussion). These discussions were used to develop a survey that provides respondents with the necessary information to complete the questionnaire, develop choice scenarios that are incentive compatible, and minimize the burden placed on respondents while collecting the necessary information.   Particular emphasis in these survey discussions was on testing for the presence of potential biases associated with poorly-designed stated preference surveys, including hypothetical bias, strategic bias, symbolic (warm glow) bias, framing effects, embedding biases, methodological misspecification, and protest responses (Mitchell and Carson 1989). Based on focus group and cognitive interview responses, EPA made various improvements to the questionnaire including making changes to ameliorate and minimize these biases in the final survey instrument.

EPA intends to implement this survey in two stages: a pretest and a main study. First, EPA will administer the pretest to a sample of 544 households using a mail survey and the Dillman Total Design Method (Dillman 2008). Assuming 92% of the sampled addresses are eligible and 30% of eligible households return the survey (see Helm 2012; Mansfield, et al. 2012; Johnston, et al. 2012), EPA estimates that this will result in 150 returned and completed pretest surveys. Households in the pretest

will be selected from each of the three geographic strata. Each selected household will be sent one version of the survey to complete (improving, constant, or declining baseline) and will be assigned that version on a random basis. Responses and preliminary findings to this pilot study will be used to inform EPA regarding the response rates and the quality of survey data. EPA will evaluate pilot responses and determine whether any changes to the survey instruments or implementation approach are needed before proceeding with the administration of the main survey.

EPA will use results from the pretest to validate the survey design. Specifically, the pretest results will be used to:

- Compare the actual and expected response rates. Based on typical mail survey response rates for surveys of this type, the expected response rate is approximately 30% (Helm 2012; Mansfield, et al. 2012; Johnston, et al. 2012).
- Assess whether demographic characteristics of the respondents are significantly different from the average demographic characteristics in the study region.
- Examine the proportion of respondents choosing the status quo. If no one is choosing the status quo, it often indicates that the cost levels are too low. Pure random selection would result in 33% of survey respondents choosing status quo. If less than 15 - 20% of responses choose the status quo in the pilot study then EPA would consider increasing the cost levels.
- Identify unusual patterns, such as the vast majority of respondents always choosing Option B. (E.g., if 2/3 of respondents (66%) choose Option B it might indicate that there is a systematic bias).
- Determine whether responses suggest that appropriate tradeoffs are being made, and that protest and other invalid responses are minimal.
- Examine response rates for individual survey questions and evaluate whether adjustments to survey questions are required to promote a higher response rate.

If required, EPA will make the appropriate adjustments to the questionnaire, sampling frame or attribute levels (e. g., increase or reduce the number of surveys mailed to households, or increase costs to households in the choice questions).

4. Collection Methods and Follow-up

4(a) Collection Methods

The survey will be administered as a mail survey. Respondents will be asked to mail the completed survey back to EPA. Several considerations justify the use of a mail survey over alternative modes (i.e., telephone and internet surveys; in-person interviews). Foremost, EPA believes respondents will face fewer obstacles in completing the mail survey compared to other collection methods. In particular, the format of key survey questions (i.e., the multi-attribute choice experiments) necessitates a visually-oriented survey, and possible substitutes to a mail survey (i.e., in-person interviews) are cost-prohibitive given the target responding sample size. Additionally, best practices for mail survey administration methods are well-established and widely accepted (Dillman 2008); EPA has designed its collection methods using these best practices and believes the mail survey will be relatively easily accessible to respondents compared to other less well-tested methods (i.e., internet surveys).

4(b) Survey Response and Follow-up

The estimated response rate for the mail survey is 30% (Helm 2012; Mansfield, et al. 2012; Johnston, et al. 2012). That is, 30% of the eligible households that are sent the mail survey are expected to return a completed survey. To obtain the highest response rate possible, EPA will follow Dillman's (2008) mail survey approach. A preview letter will be sent prior to a household receiving the survey (Attachment 5). The survey will then be sent, accompanied by a cover letter (Attachment 6), explaining the purpose of the survey, emphasizing the importance of their responses, and reminding the respondent whether they live inside or outside of the Chesapeake Bay Watershed. To improve the response rate, all of these households will receive a reminder postcard approximately one week after the initial questionnaire mailing (Attachment 7). Then, approximately three weeks after the reminder postcard, all those who have not responded will receive a second copy of the questionnaire with a revised cover letter (Attachment 8). The following week, a letter reminding them to complete the survey will be sent (Attachment 9).

5. *Analyzing and Reporting Survey Results*

5(a) Data Preparation

Since the survey will be administered as a mail survey, survey responses will be entered into an electronic database after they are returned. EPA will also clean the data to ensure that the data are entered in a consistent manner and any inconsistencies are addressed. Specifically, we will use the Double Entry data entry method for closed-ended responses. The Double Entry method consists of data being keyed twice and compared. Discrepancies are reconciled upon completion of the second entry. After all responses have been entered, the database contents will be converted into a format suitable for use with a statistical analysis software package.

5(b) Analysis

Once the survey data has been converted into a data file, it will be analyzed using statistical analysis techniques. The following section discusses the model that will be used to analyze the stated preference data from the survey.

Analysis of Stated Preference Data

The model for analysis of stated preference data is grounded in the standard random utility model of Hanemann (1984) and McConnell (1990). This model is applied extensively within stated preference research, and allows well-defined welfare measures (i.e., willingness to pay) to be derived from choice experiment models (Bennett and Blamey 2001; Louviere et al. 2000). Within the standard random utility model applied to choice experiments, hypothetical program alternatives are described in terms of attributes that focus groups reveal as relevant to respondents' utility, or well-being (Johnston et al. 1995; Adamowicz et al. 1998; Opaluch et al. 1993). One of these attributes would include a monetary cost to the respondent's household.

Applying this standard model to choices among programs to improve environmental quality in the Chesapeake Bay, a standard utility function $U_i()$ includes environmental attributes of pollution

reduction programs and the net cost of the program to the respondent. Following standard random utility theory, utility is assumed known to the respondent, but stochastic from the perspective of the researcher, such that:

(1) $U_i(\cdot) = U(\mathbf{X}_i, \mathbf{D}, Y\text{-}F_i) = v(\mathbf{X}_i, \mathbf{D}, Y\text{-}F_i) + \varepsilon_i$

where:

$\mathbf{X}_i$ =   a vector of variables describing attributes of pollution reduction program i and the baseline conditions if no further action is taken;

$\mathbf{D}$ =   a vector characterizing demographic and other attributes of the respondent.

Y =   disposable income of the respondent.

F_i =   mandatory additional cost faced by the household under program i;

$v(\cdot)$ =   a function representing the empirically estimable component of utility;

ε_i =   stochastic or unobservable component of utility, modeled as an econometric error.

Econometrically, a model of such a preference function is obtained by methods designed for limited dependent variables, because researchers only observe the respondent's choice among alternative programs, rather than observing values of $U_i(.)$ directly (Maddala 1983; Hanemann 1984). Standard random utility models are based on the probability that a respondent's utility from program i, $U_i(.)$, exceed the utility from alternative programs j, $U_j(.)$, for all potential programs $j\neq i$ considered by the respondent. In this case, the respondent's choice set of potential programs also includes maintaining the status quo. The random utility model presumes that the respondent assesses the utility that would result from each pollution reduction program i (including the "No Further Action" or status quo option), and chooses the program that provides the highest utility.

When faced with k distinct programs defined by their attributes, the respondent will choose program i if the anticipated utility from program i exceeds that of all other k-1 programs. Drawing from (1), the respondent will choose program i if:

(2) $(v(\mathbf{X}_i, \mathbf{D}, Y\text{-}F_i) + \varepsilon_i) \geq (v(\mathbf{X}_j, \mathbf{D}, Y\text{-}F_j) + \varepsilon_j) \ \forall \ j\neq i.$

If the ε_i are assumed independently and identically drawn from a type I extreme value (Gumbel) distribution, the model may be estimated as a conditional logit model, as detailed by Maddala (1983), Greene (2003). This model is most commonly used when the respondent considers more than two options in each choice set (e.g., Program A, Program B, No Further Action), and results in an econometric (empirical) estimate of the systematic component of utility $v(\cdot)$, based on observed choices among different programs. Based on this estimate, one may calculate welfare measures (willingness to pay) following the well-known methods of Hanemann (1984), as described by Freeman (2003). Following standard choice experiment methods (Adamowicz et al. 1998; Bennett and Blamey 2001), each respondent will consider questions including three potential choice options (i.e., Program A, Program B, No Further Action)—choosing the program that provides the highest utility as noted above. Following clear guidance from the literature, a "no further action" or status quo option is always included in the visible choice set, to ensure that WTP measures are well-defined (Louviere et al. 2000).

Three choice questions are included within the same survey to increase information obtained from each respondent. This is standard practice within choice experiment and dichotomous choice contingent valuation surveys (Poe et al. 1997; Layton 2000). While respondents will be instructed to consider each choice question as independent of other choice questions, it is nonetheless standard practice within the literature to allow for the potential of correlation among questions answered within a single survey by a single respondent. That is, responses provided by individual respondents may be correlated even though responses across different respondents are considered independent and identically distributed (Poe et al. 1997; Layton 2000; Train 1998).

There are a variety of approaches to such potential correlation. Models to be assessed include random effects and random parameters (mixed) discrete choice models, common in the stated preference literature (Greene 2003; McFadden and Train 2000; Poe et al. 1997; Layton 2000). Within such models, selected elements of the coefficient vector are assumed normally distributed across respondents, often with free correlation allowed among parameters (Greene 2002). If only the model intercept is assumed to include a random component, then a random effects model is estimated. If both slope and intercept parameters may vary across respondents, then a random parameters model is estimated. Such models will be estimated using standard maximum likelihood for mixed conditional logit techniques, as described by Train (1998), Greene (2002) and others. Mixed logit model performance of alternative specifications will be assessed using standard statistical measures of model fit and convergence, as detailed by Greene (2002, 2003) and Train (1998).

Advantages of Choice Experiments

Choice experiments (also called choice modeling) following the random utility model outlined above are favored by many researchers over other variants of stated preference methodology (Adamowicz et al. 1998; Bennett and Blamey 2001), and may be viewed as a "natural generalization of a binary discrete choice CV [contingent valuation]" (Bateman et al. 2002, p. 271). Advantages of choice experiments include a capacity to address choices over a wide array of potential policies, grounded in well-developed random utility theory, and the similarity of the discrete choice context to familiar referendum or voting formats (Bennett and Blamey 2001). Compared to other types of stated preference valuation, choice experiments are better able to measure the marginal value of changes in the characteristics or attributes of environmental goods, and avoid response difficulties and biases (Bateman et al. 2002). For example, choice experiments may reduce the potential for 'yea-saying' and symbolic biases (Blamey et al. 1999; Mitchell and Carson 1989), as many pairs of multi-attribute policy choices (e.g., Program A, Program B, No Further Action) will offer no clearly superior choice for a respondent wishing to express solely symbolic environmental motivations. For similar reasons, choice experiments may ameliorate protest responses (Bateman et al. 2002).

Choice experiments are well-established in the stated preference literature (Adamowicz et al. 1998; Bennett and Blamey 2001; Louviere et al. 2000) andand are commonly applied to assess values for ecological resource improvements of a type quite similar to those at issue in Chesapeake Bay TMDLs. Examples of the application of choice experiments to estimate values associated with changes in aquatic environmental quality and habitat include Hoehn et al. (2004), Johnston et al. (2002b), and Opaluch et al. (1999), among others. EPA has drawn upon these and other examples of successful choice experiment design to provide a basis for survey design in the present case.

Additionally, choice experiments permit a straightforward assessment of the impact of resource scope and scale on respondents' choices. This will enable EPA to easily conduct scope tests and other assessments of the validity of survey responses (Bateman et al. 2002, p. 296-342).

A final key advantage of choice experiments in the present application is the ability to estimate respondents' values for a wide range of different potential outcomes of Chesapeake Bay pollution reduction programs, differentiated by their attributes. The proposed choice experiment survey versions will allow respondents to choose among a wide variety of hypothetical program options, some with larger and others with smaller changes in the presented attributes (including household cost). That is,

because the survey is to be implemented as a choice experiment survey, levels of attributes in choice scenarios will vary across respondents (Louviere et al. 2000).

The ability to estimate values for a wide range of different policy outcomes is a fundamental property of the choice experiment method (Bateman et al. 2002; Louviere et al. 2000; Adamowicz et al. 1998). The experimental design (see below) will allow for survey versions showing a range of different baseline and resource improvement levels, where these levels are chosen to (almost certainly) bound the actual levels expected under pollution reduction programs. Given that there will almost certainly be some uncertainty regarding the specifics of the actual baselines and improvements, the resulting valuation estimates will allow flexibility in estimating values for a wide range of circumstances.

Comment on Survey Preparation

Following standard practice in the stated preference literature (Johnston et al. 1995; Desvousges and Smith 1988; Desvousges et al. 1984; Mitchell and Carson 1989), all survey elements and methods were subjected to extensive development and pretesting in focus groups to ameliorate the potential for survey biases (Mitchell and Carson 1989), and to ensure that respondents have a clear understanding of the policies and goods under consideration, such that informed choices may be made that reflect respondents' underlying preferences. Following the guidance of Arrow et al. (1993), Johnston et al. (1995), and Mitchell and Carson (1989), focus groups were used to ensure that respondents are aware of their budget constraints and the scope of the environmental quality improvements under consideration.

As noted above, survey development included individual cognitive interviews conducted using think-aloud or verbal protocol analyses (Schkade and Payne 1994). Individuals in these interviews completed draft survey questionnaires and provided comments and feedback about the survey format and content, their interpretations of the questions, and other issues relevant to stated preference estimation. Based on their responses, EPA made various improvements to the survey questionnaire including how the attributes are described and labeled, including an example choice question before asking respondents to complete theirs, and the appearance of the choice questions. Results from focus groups and cognitive interviews provided evidence that respondents answer the stated preference survey in ways appropriate for stated preference WTP estimation, and that respondents were evaluating trade-offs between program attributes and the household cost. The number of focus groups and cognitive interviews used in survey design, 10 focus groups and 72 cognitive interviews, exceed the numbers used

in typical applications of stated preference valuation. Moreover, EPA incorporated cognitive interviews as detailed by Kaplowicz et al. (2004).

EPA will pretest the mail survey. Making the same assumptions about eligibility and response rates as for the main survey (i.e., 92% and 30%, respectively), EPA will mail out 544 pre-test surveys (108 or 109 to each of 5 cells of the main survey, with the target number of 30 completes per cell). The main goal of the pretest is to assess whether the questionnaire is likely to produce quality data necessary estimate willingness to pay for water quality improvements in the Chesapeake Bay and lakes in the Watershed. More specifically, EPA will use pretest results to (1) assess respondents' ability to understand background information and respond to the choice questions; (2) evaluate potential for protest responses, warm glow effects, and hypothetical bias; (3) verify the expected response rate; and (4) evaluate the range and levels chosen for the cost attribute.

To test the feasibility of the non-response follow up study in the main survey, as proposed in Section 3.2(c)(iii), a small scale non-response follow up (NRFU) study will be conducted after the pre-test. EPA anticipates to mail a copy of the pilot NRFU survey to 250 households. These households will be randomly selected from the set of non-responding households from the pre-test, defined as those which did not return a completed survey response, and for which EPA did not receive a USPS delivery failure notification. EPA will send to each of the selected households a non-response follow up Priority Mail package (including the incentive). Assuming a 20% response rate, this implies that approximately fifty households will be included in the pilot NRFU study sample. The results of the pre-test NRFU will be used to refine the non-response questionnaire and to assess the expected response rate among non-respondents.

Econometric Specification

Based on prior focus groups, expert review, and attributes of the policies under consideration, EPA anticipates that five attributes will be incorporated in the vector of variables describing attributes of the pollution reduction programs (vector $\mathbf{X}_i$), in addition to the attribute characterizing unavoidable household cost F_i. These attributes will characterize improvements in water clarity (x_1), blue crab abundance (x_2), striped bass abundance (x_3), oyster abundance (x_4), and lake condition (x_5). These variables will allow respondents' choices to reveal the potential impact of Chesapeake Bay environmental quality improvements on utility.

Although the literature offers no firm guidance regarding the choice of specific functional forms for *v(·)* within choice experiment estimation, in practice, linear forms are often used (Johnston et al. 2003), with some researchers applying more flexible (e.g., quadratic) forms (Cummings et al. 1994). Standard linear forms are anticipated as the simplest form to be estimated by EPA, from which more flexible functional forms (e.g., quadratic) can be derived and compared. EPA anticipates estimating all models within the mixed logit framework outlined above. Model fit will be assessed following standard practice in the literature (e.g., Greene 2003; Maddala 1983). Since they are common in practice and theory the functional forms considered for this analysis are presented and discussed in many existing sources (e.g., Hoehn 1991, Cummings et al. 1994, Johnston et al. 1999, and Johnston et al. 2003).

For example, for each choice occasion, the respondent may choose Option A (status quo), Option B, or Option C. Assuming that the model is estimated using a standard approximation for the observable component of utility, an econometric specification of the desired model (within the overall multinomial logit model) might appear as:

$$2(\cdot) = \beta_0 + \beta_1(\text{change in water clarity}) + \beta_2(\text{change in blue crab abundance}) + \beta_3(\text{change in striped bass abundance}) + \beta_4(\text{change in oyster abundance}) + \beta_5(\text{change in lake condition}) + \beta_6(\text{Cost})$$

This sample specification allows one to estimate the relative "main effects" of program attributes on utility. Specifications such as this allow WTP to be estimated for a wide-range of potential program outcomes.

Experimental Design

Experimental design for the choice experiment surveys will follow established practices. Fractional factorial design will be used to construct choice questions with an orthogonal array of attribute levels, with questions randomly divided among distinct survey versions (Louviere et al. 2000). Based on standard choice experiment experimental design procedures (Louviere et al. 2000), the number of questions and survey versions will be determined by, among other factors: a) the number of attributes in the final experimental design and complexity of questions, b) pretests revealing the number of choice experiment questions that respondents are willing/able to answer in a single survey session, and c) the

number of attributes that may be varied within each question while maintaining respondents' ability to make appropriate neoclassical tradeoffs.

Based on the models proposed above and recommendations in the literature, EPA anticipates an experimental design that allows for an ability to estimate main effects of program attributes (Louviere et al. 2000). Choice sets (Bennett and Blamey 2001), including variable level selection, were designed by EPA based on the goal of illustrating realistic policy scenarios that "span the range over which we expect respondents to have preferences, and/or are practically achievable" (Bateman et al. 2002, p. 259), following guidance in the literature. This includes guidance with regard to the statistical implications of choice set design (Hanemann and Kanninen 1999) and the role of focus groups in developing appropriate choice sets (Bennett and Blamey 2001).

Based on these guiding principles, the following experimental design framework is proposed by EPA. A description of the statistical design is presented in Error: Reference source not found. The experimental design will allow for estimation of main effects based on a choice experiment framework. Each treatment (survey question) includes Option A (status quo) and two choice options (Option B and Option C), each characterized by five attributes and a cost variable. Hence, there is a total of eighteen attributes for each treatment (including six implicit attributes for "No Additional Action"). Based on focus groups and pretests, and guided by realistic ranges of attribute outcomes, EPA allows for three different potential levels for environmental attributes and six different levels of annual household cost for Programs A and B. It also allows for three different potential levels for environmental attributes under "No Additional Action," the first set reflecting a declining baseline, the second reflecting a constant baseline, and the third reflecting an improving baseline. The "No Additional Action" option included for each question will be characterized by a household cost of $0.

The attribute combinations can be summarized as follows:
- Water Clarity$_N$ (3 levels)
- Water Clarity$_A$, Water Clarity$_B$ (3 levels)
- Blue Crab Abundance$_N$ (3 levels)
- Blue Crab Abundance$_A$, Blue Crab Abundance$_B$ (3 levels)
- Striped Bass Abundance $_N$ (3 levels)
- Striped Bass Abundance $_A$, Striped Bass Abundance $_B$ (3 levels)
- Oyster Abundance $_N$ (3 levels)
- Oyster Abundance$_A$, Oyster Abundance$_B$ (3 levels)
- Lake Condition$_N$ (3 levels)
- Lake Condition$_A$, Lake Condition$_B$ (3 levels)

- $Cost_A, Cost_B$ (6 levels)

For a more detailed discussion of attribute levels assigned across survey versions, refer to Error: Reference source not found2.

Following standard practice, EPA will constrain the design to remove dominant/dominated pairs, where one option dominates the other in all attributes. Respondents have been found to react negatively and often protest when offered such choices. Given that such choices provide negligible statistical information compared to choices involving non-dominant/dominated pairs, they are typically avoided in choice experiment statistical designs. For example, Hensher and Barnard (1990) recommend eliminating profiles including dominating or dominated profiles, because such profiles generally provide no useful information. Following this guidance, EPA will constrain the design to eliminate such dominant/dominating pairs.

5(c)   Reporting Results

The results of the survey will be made public as part of the benefits analysis for the Chesapeake Bay TMDLs. Provided information will include summary statistics for the survey data, extensive documentation for the statistical analysis, and a detailed description of the final results. The survey data will be released only after it has been thoroughly vetted to ensure that all potentially identifying information has been removed.

REFERENCES

Adamowicz, W., Boxall, P., Williams, M., and Louviere, J. (1998). Stated preference approaches for measuring passive use values: Choice experiments and contingent valuation. *American Journal of Agricultural Economics, 80*(1), 64-75.

Anderson, E. A. (1989). Economic benefits of habitat restoration: seagrass and the Virginia hard-shell blue crab fishery. *North American Journal of Fisheries Management, 9*(2), 140-149.

Bateman, I. J., Carson, R. T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozdemiroglu, E., Pierce, D.W., Sugde, R., and Swanson, J. (2002). *Economics valuation with stated preference surveys: A manual*. Northampton, MA: Edward Elgar.

Bennett, J., and Blamey, R. (2001). *The choice modelling approach to environmental valuation*. Northampton, MA: Edward Elgar.

Bergstrom, J.C. and Ready, R.C. (2009). What Have We Learned from Over 20 Years of Farmland Amenity Valuation Research in North America? *Review of Agricultural Economics* 31(1), 21–49.

Blamey, R., and Bennett, J. (2001). Yea-saying and validation of a choice model of green product choice. In J. Bennett and R. Blamey (Eds.), **The Choice Modelling Approach to Economic Valuation.** Northampton, MA: New Horizons in Environmental Economics. pp. 178-181.

Blamey, R. K., Bennett, J. W., and Morrison, M. D. (1999). Yea-Saying in Contingent Valuation Surveys. *Land Economics* 75(1), 126-141.

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

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

Boyd, R., and Krupnick, A. (2009, Sept). **The definition and choice of environmental commodities for nonmarket valuation**. RFF DP 09-35, Resources for the Future discussion paper.

Carlson, R. E. (1977). A trophic state index for lakes. *Limnology and Oceanography, 22*(2), 361–369.

Carlson R.E., and Simpson J. (1996). A *coordinator's guide to volunteer lake monitoring methods*. Madison, WI: North American Lake Management Society.

Carson, R.T. (2012). Contingent Valuation: A Practical Alternative When Prices Aren't Available. *Journal of Economic Perspectives,* 26(4), 27-42.

Collins, J. P., and Vossler, C. A. (2009). Incentive compatibility tests of choice experiment value elicitation. *Journal of Environmental Economics and Management,* 58(2), 226-235.

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

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

Desvousges, W.H., and Smith, V.K. (1988). Focus groups and risk communication: The science of listening to data. *Risk Analysis, 8,* 479-484.

Desvousges, W.H., Smith, V.K., Brown, D.H., and Pate, D.K. (1984). *The role of focus groups in designing a contingent valuation survey to measure the benefits of hazardous waste management regulations.* Research Triangle Park, NC: Research Triangle Institute.

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

Evans, M. F., Poulos, C., and Smith, V.K. (2011). Who Counts in Evaluating the Effects of Air Pollution Policies on Households: Non-market Valuation in the Presence of Dependencies. *Journal of Environmental Economics and Management,* 62(10), 65-79.

Freeman, A.M., III. (2003). *The measurement of environmental and resource values: Theory and methods.* Washington, DC: Resources for the Future.

Greene, W.H. (2002). *NLOGIT version 3.0 reference guide.* Plainview, NY: Econometric Software, Inc.

Greene, W.H. (2003). *Econometric analysis.* 5th ed. Upper Saddle River, NJ : Prentice Hall.

Hanemann, W.M. (1984.) Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics, 66*(3), 332-41.

Hausman, Jerry. (2012). Contingent Valuation: From Dubious to Hopeless. *Journal of Economic Perspectives,* 26(4), 43-56.

Helm, E. (2012, June 05). Stated preference (sp) survey – survey methods and model results, memorandum to the section 316(b) existing facilities rule record. Retrieved from http://water.epa.gov/lawsregs/lawsguidance/cwa/316b/upload/316bmemo.pdf

Hicks, R., Kirkley, J. E., McConnell, K. E., Ryan, W., Scott, T. L., and Strand, I. (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.

Hoehn, J.P., Lupi, F., and Kaplowitz, M.D. (2004). Internet-Based Stated Choice Experiments in Ecosystem Mitigation: Methods to Control Decision Heuristics and Biases. In Proceedings of *Valuation of Ecological Benefits: Improving the Science Behind Policy Decisions*, a workshop sponsored by the US EPA National Center for Environmental Economics and the National Center for Environmental Research.

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

Johnston, R. J. (2006). Is Hypothetical Bias Universal? Validating Contingent Valuation Responses Using a Binding Public Referendum. *Journal of Environmental Economics and Management,* 52, 469-481.

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), doi:10.1029/2004WR003766.

Johnston, R.J., Schultz, E.T., Segerson, K., Besedin, E.Y., and Ramachandran, M. (2012.) Enhancing the content validity of stated preference valuation: The structure and function of ecological indicators. *Land Economics, 88*(1), 102-120.

Johnston, R.J., Magnusson, G., Mazzotta, M., and Opaluch, J.J. (2002a). Combining Economic and Ecological Indicators to Prioritize Salt Marsh Restoration Actions. *American Journal of Agricultural Economics* 84(5), 1362-1370.

Johnston, R.J., Swallow, S.K., Allen, C.W., and Smith, L.A. (2002b). Designing multidimensional environmental programs: Assessing tradeoffs and substitution in watershed management plans. *Water Resources Research, 38*(7), IV1-13.

Johnston, R.J., Swallow, S.K., Tyrrell, T.J., and Bauer, D.M. (2003). Rural amenity values and length of residency. *American Journal of Agricultural Economics, 85*(4), 1000-1015.

Johnston, R.J., Weaver, T.F., Smith, L.A., and Swallow, S.K. (1995). Contingent valuation focus groups: insights from ethnographic interview techniques. *Agricultural and Resource Economics Review, 24*(1), 56-69.

Just, R.E., Hueth, D.L., and Schmitz, A. (2004). *The welfare economics of public policy: A practical approach to project and policy evaluation*. Northampton, MA: Edward Elgar Publishing.

Kaplowicz M., Lupi, F., and Hoehn, J. (2004). Chapter 24: Multiple methods for developing and evaluating a stated-choice questionnaire to value wetlands. In Presser, Rothget, Coupter, Lesser and Martin (Eds.). *Methods for Testing and Evaluating Survey Questionnaires*. New York: John Wiley and Sons.

Kemp, W. M., Boynton, W. R., Adolf, J. E., Boesch, D. F., Boicourt, W. C., Brush, G., Cornwell, J.C., Fisher, T.R., Glibert, P.M., Hagys, J.D., Harding, L.W., Houde, E.D., Kimmel, D.G., Miller, W.D., Newell, R.I.E., Romani, M.R., Smith, E.M., and Stevenson, J. C. (2005). Eutrophication of Chesapeake Bay: historical trends and ecological interactions. *Marine Ecology Progress Series, 303*, 1-20.

Kahn, J.R., and Kemp, W.M. (1985). Economic losses associated with the degradation of an ecosystem: The case of submerged aquatic vegetation in Chesapeake Bay. *Journal of Environmental Economics and Management, 12,* 246–263.

Kling, C. L., Phaneuf, D. J., and Zhau, J. (2012). From Exxon to BP: Has Some Number Become Better than No Number. *Journal of Economic Perspectives,* 26(4), 3-26.

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

Krupnick, A., Parry, I., Walls, M., Knowles, T., and Hayes, K. (2010). *Toward a New National Energy Policy: Assessing the Options,* Resources for the Future, Washington, D.C. (November).

Krupnick, A., and Adamowicz, W.L. (2001). Supporting questions in stated choice studies. In B.J. Kanninen (Ed.), *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice.* Dordrecht, The Netherlands: Springer. pp. 53-57.

Layton, D.F. (2000.) Random coefficient models for stated preference surveys. *Journal of Environmental Economics and Management, 40*(1), 21-36.

Layton, D.F. and Brown, G. (2000). Heterogeneous Preferences Regarding Global Climate Change. *Review of Economics and Statistics,* 82(4), 616-24.

Leggett, C. and Bockstael, N. E. (2000). Evidence of the effects of water quality on residential land prices. *Journal of Environmental Economics and Management, 39,* 121-144.

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, D. and Hicks, R. (2003). The cost of stress: low dissolved oxygen and economic benefits of recreational striped bass (Morone saxatilis) fishing in the Patuxent River. *Estuaries, 26*(2A), 310-315.

Lipton, D.W., and Hicks, R. (1999). Boat Location Choice: The Role of Boating Quality and Excise Taxes. *Coastal Management,* 27(1), 81-90.

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

List, J.A. (2001). Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards. *American Economic Review,* 91(5), 1498-1507.

Louviere, J.J., Hensher, D.A., and Swait, J.D. (2000). *Stated preference methods: Analysis and application*. Cambridge, UK: Cambridge University Press.

Maddala, G.S. (1983). Limited-dependent and qualitative variables in econometrics. *Econometric Society Monographs, No.3.* Cambridge: Cambridge University Press.

Mansfield, C., Van Houtven, G., Hendershott, A., Chen, P., Porter, J., Nourani, V., and Kilambi, V. (2012). *Klamath River Basin restoration nonuse value survey, Final report*. Sacramento, CA: Prepared for the US Bureau of Reclamation. Retrieved from: http://klamathrestoration.gov/sites/klamathrestoration.gov/files/DDDDD.Printable.Klamath%20Nonuse%20Survey%20Final%20Report%202012%5B1%5D.pdf

Massey, D.M., Newbold, S.C., and Genter, B. (2006). Valuing water quality changes using a bioeconomic model of a coastal recreational fishery. *Journal of Environmental Economics and Management,* 52, 482–500.

McConnell, K.E. (1990). Models for referendum data: The structure of discrete choice models for contingent valuation. *Journal of Environmental Economics and Management, 18*(1), 19-34.

McFadden, D., and Train, K. (2000). Mixed multinomial logit models for discrete responses. *Journal of Applied Econometrics,* 15(5), 447-470.

Miller, W., Robinson, L.A., and Lawrence, R. (eds.). (2006). *Valuing Health for Regulatory Cost-Effectiveness Analysis,* Washington, DC: National Academies Press.

Mistiaen, J.A., Strand, I.E., and Lipton, D. (2003). Effects of environmental stress on blue crab (*Callinectessapidus*) harvests in Chesapeake Bay tributaries. *Estuaries, 26*, 316–322.

Mitchell, R.C., and Carson, R.T. (1989). *Using surveys to value public goods: The contingent valuation method*. Washington, D.C.: Resources for the Future.

Morgan, C., and Owens, N. (2001). Benefits of Water Quality Policies: The Chesapeake Bay, *Ecological Economics,* 39(2), 271-284,

Murphy, J.C., Allen, P.G., Stevens, T.H., and D. Weatherhead. 2005. A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation. *Environmental and Resource Economics* 30, 313-325.

Nielsen, J.S. 2011. Use of the Internet for willingness-to-pay surveys: A comparison of face-to-face and web-based interviews. *Resource and Energy Economics,* 33(1), 119-129.

NOAA. 2002. Stated Preference Methods for Environmental Management: Recreational Summer Flounder Angling in the Northeastern United States.

https://www.st.nmfs.noaa.gov/st5/RecEcon/Publications/NE_2000_Final_Report.pdf. (Accessed November 7, 2012.)

Office of Management and Budget (OMB). 2003. Circular A-4. http://www.whitehouse.gov/omb/circulars_a004_a-4 (Accessed November 7, 2012).

Office of Management and Budget (OMB). 2006. Guidance on Agency Surveys and Statistical Information Collections: Questions and Answers When Designing Surveys for Information Collections. http://www.whitehouse.gov/omb/inforeg/pmc_survey_guidance_2006.pdf

Opaluch, J.J., Grigalunas, T.A., Mazzotta, M., Johnston, R.J., and Diamantedes, J. (1999). *Recreational and resource economic values for the Peconic Estuary.* Prepared for the Peconic Estuary Program. Peace Dale, RI: Economic Analysis Inc.

Opaluch, J.J., Swallow, S.K., Weaver, T., Wessells, C., and Wichelns, D. (1993). Evaluating impacts from noxious facilities: Including public preferences in current siting mechanisms. *Journal of Environmental Economics and Management, 24*(1), 41-59.

Pagano, M., and Gauvreau, K. (2000). *Principles of biostatistics.* 2nd ed. Belmont, CA: Duxbury.

Parsons, G.R., and Thur, S.M. (2008). Valuing Changes in the Quality of Coral Reef Ecosystems: A Stated Preference Study of SCUBA Diving in the Bonaire National Marine Park. *Environmental and Resource Economics,* 40(4), 593-608.

Poe, G.L., Welsh, M.P., and Champ, P.A. (1997). Measuring the difference in mean willingness to pay when dichotomous choice contingent valuation responses are not independent. *Land Economics, 73*(2), 255-267.

Poor, P., Pessagno, K., and Paul, R. (2007). Exploring the hedonic value of ambient water quality: A local watershed-based study. *Ecological Economics, 60*, 797-806.

Schkade, D.A., and Payne, J.W. (1994). How people respond to contingent valuation questions: A verbal protocol analysis of willingness to pay for an environmental regulation." *Journal of Environmental Economics and Management,* 26, 88-109.

Train, K. (1998). Recreation Demand Models with Taste Differences Over People. *Land Economics,* 74(2), 230-239.

U.S. Bureau of Reclamation. (2012). Klamath River Basin Restoration Nonuse Value Survey. Final Report. Prepared by RTI International. RTI Project Number 0212485.001.010.

U.S. Census Bureau. (2012). 2010 Census Summary File 1. Retrieved May 31, 2012 from http://factfinder2.census.gov/.

U.S. Congress. House. (2011) Conservation, Energy, and Forestry Subcommittee of the Committee on Agrictulture. *Hearing to review the Chesapeake Bay TMDL, agricultural conservation*

practices, and their implications on national watersheds. 112th Cong., 1st sess. Washington: GPO, 2011. Print.

U.S. Department of Labor, Bureau of Labor Statistics. (2012). Table 1: Civilian workers, by major occupational and industry group. Retrieved May 2012 from: http://www.bls.gov/news.release/pdf/realer.pdf.

U.S. Environmental Protection Agency (U.S. EPA). (2008). Final Ozone NAAQS Regulatory Impact Analysis. EPA EPA-452/R-08-003. (Accessed November 7, 2012.)

U.S. EPA. (2009a). Environmental Impact and Benefits Assessment for the Final Effluent Guidelines and Standards for the Construction and Development Category. EPA-821-R-09-012. (Accessed November 7, 2012.)

U.S. EPA. (2009b). *ICR Handbook: EPA's Guide to Writing Information Collection Requests under the Paperwork Reduction Act of 1995*. Office of Environmental Information.

U.S. EPA. (2010). *Guidelines for Preparing Economic Analyses*. (EPA 240-R-10-001). U.S. EPA, Office of the Administrator, Washington, DC, December 2010.

Van Houtven, G. L. (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.

Viscusi, W.K., Huber J., and Bell, J. (2008). The Economic Value of Water Quality. *Environmental and Resource Economics*, 41(2), 169-187.

Von Haefen, R.H. (2003). Incorporating observed choice into the construction of welfare measures from random utility models. *Journal of Environmental Economics and Management, 45*, 145–165.