# Comments on the Chesapeake Bay Watershed Stated Preference Study
INTERIM DRAFT REVIEW VERSION dated Oct. 27, 2014

Nov. 26, 2014

My comments are in two parts. First, I response to the seven charge questions directly. Second, I have a list of open-ended issues and concerns. The latter is like a referee's report. Overall, I found the study design to be sound, but the empirical results left with me a series of questions (see issues and concerns below).

Response to Charge Questions

1. **Overall**

The study follows state-of-the-art state preference practices. The analytical methods are in line with the contemporary literature; it is written clearly; and it does a nice job of summarizing the literature on water quality benefits on the Chesapeake Bay. Having said all that, state preference studies are most challenged when attempting to estimate nonuse values and this study is no different than all others in this regard. I elaborate in my list of issue and concerns below.

2. **Survey Development**

The survey development process is in-line with best practices in the economic literature on stated preference studies. This includes focus groups, one-on-one interviews, pretests, careful description of resource to be valued, budget constrain reminders, clear choice format, careful sampling and so forth. There is no one right way, and researchers will debate different strategies, but this report is clearly among those in the conversation about best practices. The authors have a clear understanding of the current practices and (for the most part) have put them to use. There are a few shortcomings mentioned below.

3. **Data Description**

All of the data description is clear. The summaries are easy to follow and well done. For the most part the conclusions the authors draw are consistent with the data in the direct reporting of the results. For example, I see this consistency with the data in the reporting of the coefficient estimates and their meaning, in reporting and comparison of models, in the reporting of tests, and so forth.

4. **Data Analysis**

The specifications and methodology are in-line with literature on state preference analysis. There are a number of studies that have attempted to measure similar resources and the approach here is about the same (see for example Rob Johnston's work). The variables used make sense. The results section is fine and the interpretations make sense. The treatment of the non-response data is exceptionally good -- better than what is done in most published studies.

5. **Sensitivity Analysis**

The sensitivity analysis is in line with the current literature. I think an adding-up test (Desvousges, Mathews and Train, *Ecological Econ* 2013) and an extreme bid test would also help. The adding test is a better test than simple scope tests and trying some extreme bid amounts to see if the model's WTPs are stable to such perturbations would have made the validity checks more convincing. Perhaps the use of $500 did this but there is no reporting of the percent voting for the programs with the highest bid so it is hard to tell.

6. **Total WTP**

The total willingness to pay calculations make sense are in line with the current practices. The user and non-user treatment is also consistent with current practices.

7. **Appendices**

Balance between appendices and report itself is fine.

Concerns and Issues

Page numbers here refer to those in the INTERIM DRAFT REVIEW dated Oct. 27, 2014.

1. (Page 7, Paragraph 1) CE do not necessarily rely on a sequence of multinomial choices. One can ask a single CE question. What distinguishes CE from CV is that attributes are used to characterize the good in questions and when varied (across respondents and/or for a given respondent), can be valued.

2. (Page 10, Paragraph 4) Instead of "..we also report willingness to pay for declining and improving baseline versions of the survey", I suggest "..we also report willingness to pay for the same improvements assuming a declining and improving baseline."

3. (Page 11, Last Paragraph in Section 3.1) You write "Cost levels were chosen to ensure adequate coverage of the WTP distribution without truncation from above and were based on focus group and pilot study results." I suggest that you elaborate on what is meant by "without truncation from above." Does that mean that the percent of yes votes to the programs with the highest bid was near zero or at least very low? Pinning down the tails like this is a good idea.

4. (Page 13) Here are some interesting tests to consider (these will not be possible without further data collection): (i) treatments on the when the costs begin – next year versus in 10 years versus when the program is scheduled to begin, (ii) treatments on different baseline quantities – 25 million crabs instead of 250 million crabs, 2 million fish versus 24, and so forth. These are more serious tests of the scope and rationality. I am most concerned about how people deal with the changes in the large quantities and baseline quantities-- what does 250 million crabs and 3300 tons of oysters mean to someone who knows little about the bay? I am concerned weather you are getting 'warm glow' or real values for quantity changes here. Herein lies one of the challenges of nonuse valuation. You are not alone.

5. (Page 15) Table 1 is unclear. Not sure where the $180 through $500 costs amount are used. It seems to suggest that there are 3 fixed programs for each survey. I know that is not correct. The headers lead to the confusion. Why not just list the levels for each attribute without aligning them in columns with headers?

6. (Page 18, Paragraph 1) The decay function mentioned here is interesting and important. It does not seem to show up in your discussion of empirical results. Does WTP decline with distance from the bay? Not just being in or out of the watershed, but with distance?

7. (Pages 19 & 21 and Table 4) On page 19 you say 2200 mailings went out per stratum, but on page 21 Table 4 you say something different – that the strata are imbalanced. Clarify.

8. (Page 23, Table 6) The visitation rates (~40% over past 5 years) to the Bay seem high. Is there any outside verification of these numbers? Is raises some concern about avidity bias.

9. (Page 24) About 35% agree or strongly agree that they should not have to pay to clean up bay. Another 23% are on the fence between agree and disagree. Seems that over half of the sample is reluctant to play the WTP tradeoff game.

10. Some citations are missing from the reference list. Here are a few I stumbled across: Mansfield et al. (2011), Ariely et al (2003), and Bateman and Brouwer (2005). There may be more.

11. (Page 26) I am not convinced that the sample weights are needed in estimation. They are needed when aggregating up to WTP, but maybe not in estimate. If you assume a common model across respondents, which you do, why should there be weighting? See Deaton (2000, p. 70) for reason to prefer an un-weighted regression.

12. The survey is missing a Vossler-style consequentiality follow-up question Something like: Do you think the results of this survey will be consequential for policy? It is currently fashionable to account for those who do not find the survey consequential for policy in the validity checks. See for example Vossler et al. (2012) and Herriges et al. (2010). Nothing can be done about it now.

13. Along the same lines as above, you are missing certainty follow-up questions to deal with hypothetical bias. This has also become fashionable. See Ready et al. (2010). Nothing can be done about it now.

13. I find reading the regression results in scientific notation a bit tedious. I suggest changing the units on your covariates so that the tables are easier to read.

14. (Page 30) I have some questions about the status quo coefficient. It is relatively larger, indicating respondents are choosing some program over no action for reasons beyond the set of characteristics offered in the choice experiment. It is the size that troubles me. When I divide the SQ coefficient by the cost coefficient, that component is about $236 (=1.75/.0074). So, going from no action to some program without changing any attributes is alone worth $236. That is a sizeable step in component you would hope might be close to $0. Is this evidence of hypothetical bias? How have you handled the SQ constant in your welfare analysis? Is it included in your equation (6)? This needs to be sorted out and discussed. It would seem to overwhelm your reported values, for example in Table 17.

15. (Page 30 & Page 32, Table 12) I am surprised that the user-interactions are so weak. I would certainly have expected a positive and significant coefficient from this group and higher WTP. Perhaps distinguishing user groups by type might pick up some effects, such as angler interacted with bass population. As far as I can tell you do not have the recreation users defined this finely.

16. (Page 30) I am also surprised that the watershed-interactions are weak. I wonder if a distance variable such as distance to bay might work better. See decay function mentioned in point 6 above.

17. (Page 32, Table 12) You say units are annual. You should footnote the meaning of asterisks in your tables. Check all tables for completeness like this. Tables should "stand alone" so readers do not have to search text for meaning.

18. (Page 32) Clarity appears to be a driver of values. It would be interesting to look at the policy scenarios you considered and only change water clarity. What percent of the total value does it account for (in Table 17)? That might be important information for policy makers. General discussions around values seem to center more on the species impacts. Your results may call this into question.

19. (Page 30 and 32) Following up on the previous point, although water clarity is the most valued attribute and drives the total value, it is not significantly different from zero in Models 2 and 3 and barely significant in Model 1. This is problematic.

20. (Page 33, Table 13) I would dispense with the ethnicity and education interactions. These are not really important and are distracting. I would also put the income term directly in Models 1, 2 and 3. In this way models 4, 5 and 6 can be dropped. Then, going forward in the report, stay with Model 3. I don't think you need to show us WTP for Models 1 and 2, just say they are close and move on. This will tighten the report.

21. (Table 13) It also disappointing that the income terms does not "work", but this is common in nonuse value estimation.

22. (Page 39) The incremental changes realized for policy are smaller than the increments used in the analytical model. For example, the smallest change people see in the survey for oysters is 500 tons, but one of the policies looks at a 6-ton change. Another example is bass. The smallest change measured in the survey is 15 million with some near 100 million but the policy changes are smaller, like 0.3 million and 1 million. A closer correspondence here would be better.

23. (Page 44) I think an adding-up test (Desvousges, Mathews and Train, *Ecological Econ* 2013) and an extreme-bid test would be useful validity checks. The adding test is a better test than simple scope and trying some extreme bid amounts to see if the model's WTPs are stable to this perturbation would have made the validity checks more convincing.

24. There is not much discussion of the standard deviation estimate on the status quo term. I am assuming this was estimated as a mixed logit, but find little analysis over that aspect of the model. It would useful to add some discussion.

25. (Page 45, Table 21) Over the variables clarity, bass, crab, and oysters, there are 12 possible coefficient estimates (3 models, 4 variables). Although the signs are 'correct' in 11 of the 12 cases, the coefficients are statistically significantly different from zero in only 4 of the 12 cases. I would call this weakly passing scope. The sample sizes here (>1000) are certainly sufficient for finding statistically significant effects if they exist. The coefficient estimates are also highly unstable across these models.

26. Check table numbering around page 45. It appears to go ... 21, 23, 22, 23, 24 ... .

27. (Page 47, Table 23) The stepwise model is informative. It seems to say for clarity, bass, and crab that a "full step" is beneficial but a "partial or first step" is not. The oyster results seems to say less is better than more (at least after you reach a certain level), and lakes fit a nice stepwise story of nonlinear but increasing effects with size. You offer this as a scope test, but why not use this as your model? Linearization of the model masks the nonlinear effects shown here. I guess the other thing that concerns me is that the policy scenarios are largely over small steps or something closer to or less than the partial step. By using a linear model you are, in effect, using the effects for larger improvements to extrapolate to smaller steps in improvements. This needs to be reconciled. (Comment 22 above.)

28. (Page 51, Paragraph 2) Reference to tables 27 and 28 are incorrect.

29. (Page 54) It is not clear why the increasing baseline was only administered to Bay stratum.

30. (Page 55) It is not clear to me why the status quo term (SQC) is dropped in the changing baseline cases. It should be feasible. I do not understand the co-linearity mentioned here. The status quo is not really "declining" or "increasing" in either case, there is simply a different (lower or higher)

reference level in the choice question. I may be missing something. Some clarification would help.

31. (Page 55) It would be interesting to see the WTP for the same policy targets across the differing baselines to see if WTP increases with a decreasing baseline and decreases with an increasing baseline as expected. This actually a form of a scope test.

References

Deaton, Angus (2000). *The Analysis of Household Surveys: A Microeconometric Approach to Develop Policy*. Baltimore: Johns Hopkins University Press.

Herriges, J., Kling, C., Liu, C. & Tobias, J. (2010). "What are the Consequences of Consequentiality?" *Journal of Environmental Economics and Management* 59(1):67-81.

Ready, R., Champ, P. & Lawton, J. (2010). "Using Respondent Uncertainty to Mitigate Hypothetical Bias in a Stated Choice Experiment". *Land Economics* 86(2):363-381.

Vossler, C., Doyon, M. & Rondeau, D. (2012). "Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments." *American Economic Journal: Microeconomics* 4(4):145-71.