Preliminary Individual Comments from Drs. Kevin Boyle and Richard Ready

(March 6, 2016)

Dr. Kevin Boyle

1a. Evidence of validity for stated preference studies: The SAB noted in its earlier advisory report (U.S. EPA Science Advisory Board 2011) that each selected stated preference study "should provide evidence that it yields valid estimates" (page 16). The SAB did not, however, specify how validity should be assessed. In applying this criteria, EPA included studies and estimates that passed a weak scope test or provided other evidence of validity (e.g., a positive coefficient on the risk variable as in the appendix for Viscusi, Huber and Bell 2014) as explained in Appendix B of the White Paper. Please comment on whether the methods EPA used in the White Paper to assess the validity of studies and estimates are appropriate and scientifically sound.

I think that it is important to be careful with the concept of validity. There are three concepts of validity – content, construct and criterion (Carmines and Zeller, 1979). Content has to do with the use of established procedures to implement a method (e.g., Guidelines for Preparing Economic Analyses), construct has to do with the testing of specific procedures (e.g., procedural invariance, convergent validity, tests of scope, etc.), and criterion validity has to do with comparison of empirical outcomes against a presumed truth (e.g., comparisons with cash transactions).

There is no perfect study and no absolute test of validity. Follow all of the recommended procedures does not ensure that a study will not produce biased results, only implies that the likelihood of unbiased results is enhanced or, perhaps, that bias is minimized. Satisfying construct validity only establishes the credibility of a specific procedure. Criterion validity is the strongest concept of validity, but the outcome is only as credible as the credibility of the criterion as the measure of the presumed truth. Validity of any specific study, or value estimate, is a matter of the weight of evidence.

` The assessment of the validity of stated-preference VSL estimates falls under the concept of construct validity. Consideration of question-ordering effects is a test or procedural invariance; question order should not affect value estimates and if this holds, this is evidence of construct validity. It is logical to assume that people would pay more for a larger risk reduction and for a smaller risk reduction, and for a given risk reduction, the probability that any person would pay for the risk reduction decreases as the cost increases.

These are appropriate considerations to investigate the validity of value estimates. However, they are not absolute evidence of validity or invalidity. Violation of procedural, due to a question ordering effect, does not establish whether one or both value estimates are biased. While establishing a price effect or a scope effect provides evidence of validity, but the absence of these effects does not refute validity as these are investigator imposed relationships that may or may not be consistent with preferences.

Failure of any specific validity investigation does not mean a value estimate is invalid and the same implication holds for any affirmative validity investigation. Thus, consideration of validity should

consider as many relevant features of a study and the procedures used to estimate a value as possible to bring the weight of evidence to support any validity judgement.

Other study features that could be considered to investigate validity include but are not limited to the type of valuation question (e.g., referendum) and consequentiality of the scenario design and choice. Other study features may not have clear insights on validity for the literature, such as the unit and time frame of payments, and these considerations could be investigated through robustness analyses.

Having said that there are other study features that can be considered to investigate validity, studies that show evidence of invalidity can still help inform any meta-analysis of the empirical literature. Thus, a broader consideration of validity can inform the weight of evidence validity assessment and enhance the credibility of the meta-analysis. At the same time casting a wide net, even if there is some evidence of invalidity, can potentially enhance the insights gained from a meta-analysis.

A final comment is that the evidence of validity discussions are interspersed with the selection criteria for including observations in the meta-analysis. This makes the evidence of validity convoluted in the reporting and could be enhanced by separating these two selection criteria in the reporting. Such a format change would allow clear comparisons of validity evidence across studies. In addition, the validity of the meta-analysis weighs heavily on the observation selection and this separation of the reporting could enhance the credibility of the meta-analysis as well.

It is worth noting that excluded studies and selected observations all appear to be based on the selection criteria for the meta-analysis. The validity evidence only appears to be presented as auxiliary information.

Carmines, E.G. and Zeller, R.A., 1979. *Reliability and validity assessment* (Vol. 17). Sage publications.

U.S. EPA. 2010. *Guidelines for Preparing Economic Analyses*. National Center for Environmental Economics, Office of Policy, U.S. Environmental Protection Agency (updated May 2014).

1b. Construct of the risk variable in hedonic wage studies: The SAB noted in its earlier advisory that the EPA should "Eliminate any study that relies on risk measures constructed at the industry level only (not by occupation within an industry)" (U.S. EPA Science Advisory Board 2011, page 18). It is not clear whether the SAB's parenthetical addition was meant as an example or as a directive. Only four studies constructed the risk variable by occupation and industry and met other selection criteria. In applying this criteria EPA included studies and estimates where the risk measure is differentiated by industry and at least one other characteristic (e.g., occupation, gender, age). Please comment on whether the hedonic wage studies included in the White Paper constructed the risk variable in a manner appropriate for use in the meta-analysis.

It would seem that the use of industry-level risk measures would obscure important variations of risk within an industry that could inform a meta-analysis. However, this may or may not be the case given the type of position and risk evaluated.

It is possible that the studies that do not meet this criterion could still help to inform the meta-analysis. The concern, however, even for the included studies, is whether all studies are measuring a value for a common risk reduction, i.e., is the dependent variable all apples or a mixture of apples, oranges, and other fruit? Here is an area where robustness analyses are warranted.

An auxiliary question is why validity of the hedonic wage estimates is not considered as done for staed-preference estimates.

Dr. Richard Ready

Charge Questions for SAB-EEAC Review of an EPA White Paper: "Valuing mortality risk for environmental policy: a meta-analytic approach" and Technical Memorandum: "Income Elasticity of VSL"

February 2016

White Paper: Meta-analysis dataset

The White Paper assembles a database of stated preference and hedonic wage estimates of the value of statistical life (VSL) and, where possible, their standard errors. Criteria for inclusion in the database are based on recommendations from the SAB-EEAC (U.S. EPA Science Advisory Board 2011) (see section 4.4, page 13-20). EPA requests comments on whether the selection criteria previously recommended by the SAB-EEAC were appropriately interpreted and applied both for selecting studies to include in the meta-analysis and for selecting estimates within studies. **In answering questions 1(a) – 1(c), in addition to responding to the specific questions, please comment, in general, on whether the selection criteria previously recommended by the SAB-EEAC have been appropriately interpreted and applied in the White Paper.**

1a. Evidence of validity for stated preference studies: The SAB noted in its earlier advisory report (U.S. EPA Science Advisory Board 2011) that each selected stated preference study "should provide evidence that it yields valid estimates" (page 16). The SAB did not, however, specify how validity should be assessed. In applying this criteria, EPA included studies and estimates that passed a weak scope test or provided other evidence of validity (e.g., a positive coefficient on the risk variable as in the appendix for Viscusi, Huber and Bell 2014) as explained in Appendix B of the White Paper. Please comment on whether the methods EPA used in the White Paper to assess the validity of studies and estimates are appropriate and scientifically sound.

Response: The criteria for inclusion are reasonable, with the exceptions of the requirement for scope sensitivity and the requirement for peer review. The required degree of scope sensitivity for stated preference studies is not described. It is not enough to observe that WTP is higher to avoid larger risks. The relationship should be roughly proportional. Peer review is neither necessary nor sufficient to establish reliability. Plenty of methods-focused studies are published that should not be used for policy analysis, and plenty of high-quality, policy-relevant studies remain unpublished because they did not include a methodological twist. Validity should be evaluated on a case-by-case basis.

1b. Construct of the risk variable in hedonic wage studies: The SAB noted in its earlier advisory that the EPA should "Eliminate any study that relies on risk measures constructed at the industry level only (not by occupation within an industry)" (U.S. EPA Science Advisory Board 2011, page 18). It is not clear whether the SAB's parenthetical addition was meant as an example or as a directive. Only four studies constructed the risk variable by occupation and industry and met other selection criteria. In applying this criteria EPA included studies and estimates where the risk measure is

differentiated by industry and at least one other characteristic (e.g., occupation, gender, age). Please comment on whether the hedonic wage studies included in the White Paper constructed the risk variable in a manner appropriate for use in the meta-analysis.

Response: The concern with using industry-level risk estimates is that risk varies tremendously within an industry, but should be more homogeneous within a specific occupation within an industry. Segregating an industry by other dimensions (gender, age, etc) will not much to solve the heterogeneity problem.

1c. Estimates for immediate risk reductions: To estimate the average value of the marginal willingness to pay for reduced risk of immediate death, the EPA selected estimates from the Stated Preference literature that are most closely comparable to the accidental deaths from the hedonic wage literature. The EPA made several judgement calls in determining the appropriate estimates to use from the stated preference literature. Specifically, Viscusi, Huber and Bell (2014) estimate reductions in risk of bladder cancer that will occur in 10 years. The authors discount the estimates to derive a comparable estimate for an immediate risk reduction. Alberini, et al. (2004) estimate a willingness to pay for an annual reduction in risk over 10 years. We include estimates from both of these studies in the meta-analysis. Please comment on whether appropriate estimates from the stated preference literature were used in the White Paper to estimate the marginal willingness to pay for reduced risk of immediate death.

Response: The VHB study was included based on an online appendix that purports to demonstrate sensitivity to risk levels. However, I am unconvinced that the analysis included in that appendix demonstrates scope sensitivity. In that appendix, the authors estimate a RUM of the form

$$U = a*risk + B*cost + error$$

In the "with filter" case, risk is set equal to either 0, 1, 2 or 3/100,000, and cost varies across respondents from $50 to $400. In the "without filter" option, risk is set at 2 or 4/100,000, and cost is 0. The authors estimate a probit regression with change in risk and cost as the two explanatory variables, and find both variables to be significant in the regression. Note that the authors' model assumes risk sensitivity. Within their model, the only way for respondents to be risk-insensitive is if a=0, in which case almost no one would choose to install filters (only those with large positive errors for the filter option). To test for risk sensitivity, the authors should have included an alternative specific constant for the "with filter" option. Had they done so, and still observed a significant risk parameter, then we would conclude that the responses show (weak, at least) scope sensitivity.

2. Please comment on whether relevant empirical studies in the stated preference and hedonic wage literatures are adequately captured in the White Paper. If additional studies should be included in the white Paper please provide citations.

3. Some estimates in the meta-analysis dataset in the White Paper are constructed by weighting subpopulation-specific estimates within a study in order to approximate an estimate for the

general population. The specific weights used are described in Appendix B of the White Paper. Please comment on whether the population-weighting approach used in the White Paper is appropriate and scientifically sound.

4. In some cases EPA estimated standard errors in the White Paper using information within studies or provided by the study authors, as described in Appendix B. Please comment on whether the methods used in the White Paper to estimate standard errors when such information was not readily available are appropriate and scientifically sound.

White Paper: Analysis

Section 4 of the White Paper describes methods used to estimate representative VSL estimates from the meta-analysis dataset and presents results.

5. Please comment on whether the methodology used in the White Paper to analyze the data represents an appropriate and scientifically sound application of meta-analytic methods to derive generally applicable VSL estimates for environmental policy analysis.

Response: The non-parametric and parametric approaches are appropriate and sound. The parametric approach is not really much more restrictive than the non-parametric approach in terms of the assumptions required about the structure of the errors, and allows consideration of a time trend, which is valuable. For this reason, I prefer the parametric approach to the nonparametric approach.

6. The White Paper classifies estimates into independent samples, also called groups, as described in Section 4. Estimates from some hedonic wage studies that use the same or very similar worker samples are grouped together for the analysis. Similarly, some of the stated preference estimates using the same sample are grouped together. Please comment on whether this methodology represents an appropriate and scientifically sound approach for accounting for potential correlation of results that rely on the same underlying data.

Response: This approach is appropriate and sound, with the exception of situations where both mean and median estimates are obtained from the same sample for the same good. These are not independent estimates. They are simply different statistics calculated from one estimated distribution. In general, median estimates of VSL are not appropriate for RIAs. The meta-analysis should be restricted to mean estimates.

7. Section 4.1 of the White Paper presents an expression that characterizes optimal weights that account for sampling and non-sampling errors, a framework that guides EPA's approach. Please comment on whether this is an appropriate and scientifically sound approach for addressing sampling and non-sampling errors.

8. The analysis in the White Paper adopts both non-parametric and parametric approaches (sections 4.1 and 4.2, respectively). Please comment on whether these approaches span a reasonable range of appropriate, scientifically sound, and defensible approaches to estimating a broadly applicable VSL for environmental policy and whether there are other methods that are more appropriate than those used in the White Paper.

Response: See comment above.

White Paper: Results

9. The White Paper presents estimates using parametric and non-parametric models, pooled across stated preference and hedonic wage studies as well as balanced (i.e., equal weight to each study type), and weighted using different approaches. Of the range of estimates presented (see Section 4) the White Paper proposes the use of estimates from the following models:

    - Non-parametric model, balanced, mean of study mean
    - Parametric, balanced

    Please comment on whether these proposed estimates represent reasonable and scientifically sound conclusions from the analyses in the White Paper and whether there is a different set (or sets) of results that are preferable based on the data and analysis in the White Paper.

Response: The White Paper does not provide a clear argument why the balanced estimates are preferred to the pooled estimates. This has the effect of placing more weight on each SP study.

10. The results section of the White Paper concludes with an influence analysis. Please comment on whether this analysis is a reasonable way to characterize the influence of individual studies on the estimated VSLs, whether the results of the influence analysis suggest any changes or modifications to the estimation approach, and whether it is important to include an influence analysis.

The influence analysis is interesting, but it is not clear how it should be used. A large influence should not necessarily signal that a study is suspect.

Establishing a Protocol for Future Revisions:

11. In the previous SAB advisory report (USEPA Science Advisory Board 2011), the SAB endorsed the idea of establishing a standardized protocol and regular schedule for future updates to the Agency's mortality risk valuation estimates. Please comment on relevant statistical criteria for the inclusion of additional eligible estimates and/or the exclusion of older estimates that could help inform the development of a standardized protocol for future updates and the timing or frequency of those updates.

As time progresses, and some of the studies included in the current meta-analysis age, the EPA should consider whether they adequately reflect current risk preferences. The meta-analysis demonstrates that WTP for risk reductions has increased over time, at rates larger than what would be expected from changes in wealth alone. This means that new VSL estimates will continuously be needed, and older estimates should at some point be retired or downweighted. The criteria for inclusion used in the current White Paper are reasonable, but the bar should be raised when it comes to evaluating scope sensitivity to risk in stated preference studies – having an upward slope is not enough, WTP should be roughly proportional to the risk reduction. Improvements in risk communication and survey delivery technology should allow better engagement by respondents.

12. In its 2011 report the SAB-EEAC recommended "…EPA work toward developing a set of estimates…for policy-relevant cases characterized by risk…" (U.S. EPA Science Advisory Board 2011, pp. 10). Among the studies that meet the selection criteria in the current White Paper, three stated preference studies provide values for reductions in risks of cancer (i.e., Hammitt and Haninger 2010, Chestnut, Rowe, and Breffle 2012, and Viscusi, Huber and Bell 2014). Only two of those studies (Hammitt and Haninger 2010 and Chestnut, Rowe, and Breffle 2012) allow for a within study comparison of values for cancer and non-cancer risk reductions. However, EPA could augment the literature by modifying the selection criteria to include studies from other countries or from the grey literature, and/or using other methods (e.g., risk-risk studies). Please comment on whether, and if so how, selection criteria for identifying studies for estimating a cancer differential should differ from those used in the current White Paper. Does the literature support a non-zero cancer differential?

Response: When considering whether a cancer differential should be applied, I see three main issues. First, why should cancer be singled out? Other mortality risks associated with exposure to environmental pollutants and toxins also carry some degree of dread (such as congestive heart failure, COPD, heart attack). Second, some of the motivation for a cancer differential (or a differential for any other disease-related death caused by environmental exposure) is that, in contrast to accidental deaths, disease-caused deaths include some period of pain and suffering. In other words, a death from cancer is viewed as two things, a period of pain and suffering followed by an early death. If a disease premium is applied, the EPA needs to be careful to avoid double counting, whereby the same case of illness generates both morbidity (valued in the form of restricted activity days, symptom days, etc) and mortality (valued using a VSL). If the VSL is adjusted upward to reflect the morbidity, then the morbidity must not also be valued. A third issue is whether there is empirical evidence that a cancer (or other disease) premium exists. As noted, the two within-study comparisons done to date have not found a premium. Specifically, they found that the VSL for a cancer was no different from VSL for other diseases with similar morbidity. The VHB study compares a stated preference VSL for cancer to VSL estimates for (mostly accidental) deaths from multiple other sources. The difference found in this between-study comparison is not compelling, and is no larger than differences found within the VHB study when comparing different estimates of the VSL for cancer obtained by analyzing the same dataset in different ways. The EPA should continue to explore this area, being careful to distinguish between WTP to avoid deaths from different causes vs WTP to avoid deaths plus morbidity from different causes. In the

meantime, there is not sufficient evidence to motivate a cancer premium, or a premium for any other environment-related mortality. The EPA could look to studies conducted in other countries and studies in the gray literature that include within-study comparisons of VSLs for different mortality risks, but those studies should be used for background information only. Differences in culture and healthcare systems make inter-country benefit transfer dangerous, even when it is only the relative premiums that are being transferred, not the values themselves.

Technical Memorandum: Income elasticity

13. The EPA document *Technical Memorandum: Income Elasticity* presents a summary of the recent income elasticity literature based on a review presented in Robinson and Hammitt (2015). Please comment on whether Robinson and Hammitt (2015) and the EPA Technical Memorandum provide an appropriate and scientifically sound summary of the income elasticity of VSL (IEVSL) and income elasticity of non-fatal health effects literatures. If there are additional relevant empirical studies that should also be included in the summary, please provide citations.

Response: Care should be taken when using Income elasticities estimated from cross section analysis to calculate changes in mean VSL across the population resulting from population-wide increases in income. In most of the studies included, the income elasticity of VSL is estimated based on variation in VSL across individuals, i.e. at a given point in time, lower-income individuals tend to have lower VSL than higher income individuals. This is the source of variation from which the income elasticity is estimated. There are two issues that should be recognized. First, it is not necessarily the case that cross-sectional variation in income will have the same impact on VSL as temporal variation. The White Paper analysis of VSL estimates suggests that VSL has been increasing faster than would be expected from income growth, even using a relatively high estimate of income elasticity. A population-level income elasticity could be estimated using meta-analysis incorporating studies conducted at different times, though I wonder whether there is enough variation to reliably identify the elasticity that way and distinguish it from temporal changes in preferences. Second, even if every individual in the population shares a common income elasticity for VSL that is constant over time, the distribution of income changes over the population matters. If every individual's income rises by 1%, then mean VSL will increase by the common income elasticity. If, however, income gains are concentrated in upper income households, and risk reductions are a luxury good, then mean VSL will tend to increase by more than the common income elasticity. This effect would be further exacerbated if the income elasticity of VSL tends to be higher for higher income households.

14. Several reported mean income elasticity estimates from stated preference studies are quite low, sometimes even zero. The "balanced" approach in the EPA Technical Memorandum does not include reported mean estimates of zero, but does include very low reported mean estimates (e.g., 0.1). Please comment on whether this an appropriate and scientifically sound choice. How should very low, non-zero, mean reported income elasticity results be addressed in the analysis?

Response: All of these low estimates of the income elasticity are from cross-sectional analyses. For the reasons stated above, I have less confidence that cross-sectional analyses will provide useful information for tracking changes in the VSL over time. Within cross-sectional analyses, there is not a strong argument that studies should be included or excluded based on the estimated size of the elasticity. If we

have strong a priori expectation, based on theory, that the income elasticity should be of a certain size (say, 1), then we should just impose that, rather than select studies to support our expectations.

15. Please comment on whether the selection criteria applied by Robinson and Hammitt (2015) are clearly enumerated, appropriate, and scientifically sound and whether the additional inclusion of Viscusi, Huber, and Bell (2014) in the Technical Memorandum is appropriate based on results reported in the study's on-line appendix (attached).

Response: The criteria are sound, but I do not believe that the VHB paper has sufficiently demonstrated scope sensitivity. See my comments to item 1c.

16. Given the relatively limited number of studies upon which to draw for estimating the income elasticity of VSL, the EPA Technical Memorandum describes two alternatives for arriving at a central IEVSL estimate and range for use in environmental policy analysis. Of these alternatives which is the most appropriate and scientifically sound? Please provide the rationale for your choice. Would it be appropriate to consider using the alternative as a sensitivity or uncertainty characterization?

Response: Because it does a better job capturing temporal variation in income, and its effect on the VSL, I prefer Option 1.

17. As described in Robinson and Hammitt (2015), there are limited data on income elasticity of non-fatal health effects. As a result the Technical Memorandum recommends using the IEVSL to estimate income elasticity for the value of these non-fatal health risks. Please comment on whether this represents an appropriate and scientifically sound approach given the available data.

Response: This approach is probably the best approach for now, but the EPA should encourage studies that explicitly look at the income elasticity of WTP to avoid morbidity, taking into account the comments I made regarding cross-sectional vs temporal variation in income.