

**Compilation of Preliminary Comments from
Individual Committee Members**

(as of March 3, 2016)

<i>Dr. Sylvia Brandt</i>	2
<i>Dr. Mary Evans</i>	5
<i>Dr. Wayne Gray</i>	13
<i>Dr. F. Reed Johnson</i>	15
<i>Dr. Matthew Kotchen</i>	21
<i>Dr. Matthew Neidell</i>	25
<i>Dr. James Opaluch</i>	28
<i>Dr. Andrew Plantinga</i>	38
<i>Dr. Kerry Smith</i>	41
<i>Dr. George Van Houtven</i>	63
<i>Dr. JunJie Wu</i>	72

Dr. Sylvia Brandt

Charge question #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.

Preliminary Comments by Andrew Plantinga and Sylvia Brandt

Assessment of methods

The question about whether the methods “span a reasonable range of appropriate, scientifically sound, and defensible approaches” could have been answered, in part, by the authors of the white paper. There are no citations given for the estimators in 1-5 on pages 22-23. While the methods are clearly presented, they could be better justified in terms of the particular application—finding the central tendency of VSL estimates from studies that in most cases report multiple estimates.

Equation (1) is the object of interest and all five approaches represent alternative schemes for defining the weights w_{ij} . Estimator 1 and 2 are straightforward approaches but do not make use of information in the studies on sample size or sampling error variance. Estimator 1 (simple mean) is potentially problematic because it gives the most weight to papers with the most estimates. It is not clear why there is so much variation among studies in the number of reported estimates. Are the papers with a lot of estimates reporting robustness checks, demonstrating differences in estimates resulting from specification changes, or something else? Either way, it is not clear that multiple estimates from a single paper should be treated as if they come from separate papers. Estimator 2 (mean of group means) has the potential advantage that it gives the same weight to each paper.

The sample size weighted mean (estimator 3) uses information on sample size and appears to be a standard approach (it is described in the text on meta-analysis by Hunter and Schmidt (2004)). Likewise, the sampling error variance weighted mean uses information on the sampling error variance. It is described in another text on meta-analysis (Hedges and Olkin 1985) and implemented in a recent meta-analysis by Hsiang et al. (2013).

Estimator 5 involves directly estimating the group and observation-level non-sampling errors. The estimate of the individual non-sampling variance requires that the total variance of each estimate, $\text{var}(y_{ij})$, be estimated from within-group variation (equation 5). There are only four studies that provide more than four observations and so these estimates are likely to be very imprecise. The authors take a weighted average of these estimates (equation 8) to obtain a single

individual-level non-sampling variance estimate. Other than practicality, it is hard to see a justification for this assumption. Our assessment is that this approach makes sense if there are a large number of estimates from each study.

The parametric estimation includes study characteristics: year data collected, SP or HW, median. It is not clear what the median dummy captures, although our guess is that it signifies that a median estimate is reported. The included controls should be justified and explained. For example, what is the year variable intended to capture? The specification seems fairly sparse compared to other meta-analysis regression analyses. Are there other controls that could be included? Some possibilities include: whether the given data set was used, whether the studies focused on the same or a similar population, whether a given researcher was a co-author.

Alternative tests

To address the issue, mentioned above, of equal weight given to multiple estimates from a paper, we suggest exploring whether the mean of group means estimator could be blended with estimator 3 and with estimator 4 which take into account information on sample size and sampling error variance. In the first case, estimator 2 could be modified so that each group mean is weighted by the average sample size for study j .

An alternative approach to the estimates in 1-5 is to compute the median across studies and see if this value lies within the confidence interval for the individual estimates. This approach is robust to outliers, but fails to make use of other information.

These estimates assume no correlation in the VSL estimates across studies. Following Hsiang et al. (2013), one could assume positive correlation across such studies and investigate if there is a consequential effect on the variance of the average estimate.

Can the parametric model (11) be estimated with weighted least squares? It appears to have the same structure as a standard random effects model. If not, the reason why the simulated likelihood approach, which requires additional structural assumptions, is needed should be explained.

The authors could conduct a publication bias test, following Card and Krueger (1995). The basic idea is to look at t-stats on the estimate of interest and see if it is positively related to the sample size (or degrees of freedom). If researchers are selectively reporting results, then there should be no or a negative relationship.

References

Hsiang, S.M., Burke, M., and E. Miguel. 2013. Quantifying the Influence of Climate on Human Conflict. *Science* 341 (2013): 1235367.

3/3/16 Preliminary draft comments from individual members of the SAB Environmental Economics Advisory Committee. These comments do not represent consensus SAB advice or EPA policy.
DO NOT CITE OR QUOTE

L. V. Hedges, I. Olkin, *Statistical Method for Meta-Analysis* (Academic Press, 1985).

D. Card, A. B. Krueger, Time-series minimum-wage studies: A meta-analysis. *Am. Econ. Rev.* 85, 238 (1995).

Dr. Mary Evans

Mary F. Evans (Claremont McKenna College)

Draft responses to 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”

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:

This criterion seems an appropriate interpretation of the SAB recommendation. To the extent that some stated preference studies could be identified as providing significantly more evidence of validity than other studies (i.e., perhaps as in the IEVSL paper), a sensitivity analysis could be used to explore the implications of excluding estimates from the latter type of studies.

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:

Given the paucity of studies using fatality risk measures that vary by industry and occupation, EPA's inclusion of studies with risk measures that vary by industry and another characteristic seems reasonable. However, it would be helpful to report as a robustness check the sensitivity of the results to the exclusion of studies that measured risk by industry and something other than occupation (i.e., restrict the hedonic wage estimates to those reported in Viscusi (2004), Kneisner and Viscusi (2005), Scotton and Taylor (2011), and Scotton (2013)).

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.

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.

Response:

An additional hedonic wage study to consider is Viscusi and Gentry (JRU, 2015) (a subset of results uses the full sample rather than limiting attention to transportation-related fatalities, the primary focus of the paper).

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.

Response:

Weighting the VSL estimates by representation in the population as has been proposed is reasonable. I'm uncertain as to whether or not this is appropriate for the standard errors (given the issue I discuss below).

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.

Response: The white paper fails to provide detailed information about how the standard error of the VSL is calculated in situations where one is not reported in the original study. In such cases, the white paper notes that the standard error of the VSL is calculated "based on the standard error of the risk coefficient alone". The exact formula used is not provided.

If the study provides the standard deviation of the wage and the sample size (in addition to the estimated risk coefficient, its standard error, and the sample mean wage), then there is sufficient information to accurately calculate the standard error of the VSL. I describe how in what follows.

Let w denote the weekly wage.¹ w is a random variable with unknown population mean and standard deviation. Let μ_w and σ_w denote the sample mean and standard deviation, respectively. μ_w is an unbiased estimate of the population mean wage. The standard error is calculated as follows:

$$se_w = \frac{\sigma_w}{\sqrt{N}}$$

where N denotes the sample size.

Let $\hat{\beta}$ represent the estimated coefficient on the occupational fatality risk variable (i.e., the estimate of the true parameter β) and $se_{\hat{\beta}}$ its standard error. Assume risk is measured as the number of fatalities each year per 10,000 workers in the occupation-industry category.

Assuming a log linear specification and that each worker works 50 weeks per year (i.e., treating this as a constant), the estimated VSL is then given by:

$$\widehat{VSL} = 10,000(50)(\hat{\beta})(\mu_w) = 50,000\hat{\beta}\mu_w$$

The standard error of the VSL is given by:²

¹ In cases where wage is reported as hourly, the formula differs somewhat from that derived in what follows.

² The calculation assumes β and w are independent random variables makes use of the following formulas. The variance of the product of a constant a and a random variable X is given by $a^2 var(X)$. The variance of the product of two independent random variables X and Y is given by $var(X)var(Y) + var(X)[E(Y)]^2 + var(Y)[E(X)]^2$.

$$se_{VSL} = \sqrt{\{(50,000)^2 [(se_{\beta})^2 (se_w)^2 + (se_{\beta})^2 (\mu_w)^2 + (se_w)^2 (\hat{\beta})^2]\}}$$

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.
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:

While the groups are generally referenced in the white paper, I found it difficult to identify the exact composition of the groups from the main text and the appendices. It would be helpful to add a column to Table 6 of the white paper that provided this information. It might also be reasonable to consider how alternative definitions of the groups affect the overall results.

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.

Response:

A few comments:

1. Please clarify the distinction between the observation-level non-sampling error and the observation-level sampling error.
2. Should σ_{η}^2 have an i subscript (i.e., should the variance be group-specific)?
3. I've been unable to replicate expression (4). Please provide more detail on the derivation of this expression.

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:

A few questions/comments:

1. Regarding the parametric analysis, previous recommendations from the SAB suggested not adjusting VSL estimates for income growth but instead exploring how VSL “varies with (a) the time period to which the data pertain and (b) the average sample income as part of the meta-analysis.” The parametric model includes controls for study year but the primary estimates are adjusted using an assumed income elasticity of 0.7 (with alternative values considered in an appendix). Are the parametric results comparable when the estimates are unadjusted for income growth but controls for year and mean sample income included (consistent with the previous SAB recommendation)?
2. The difference between mean VSL and median VSL estimates should not be cited as a justification for preferring the mean estimates to the median estimates (p. 31). Both mean and median estimates provide useful information about the VSL distribution.

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.

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.

Response:

In previous responses I suggest alternative sensitivity analyses that could complement the influence analysis currently included in the paper. It might be helpful to think about what results in the influence analysis (e.g., a threshold) would result in a study's exclusion.

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.
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:

In my opinion, the literature has not advanced to a point at which one can decisively make conclusions about a cancer differential. I'm curious why the Cameron and DeShazo (2013) study is excluded from the list of studies referenced in the charge question as it would seem to meet the selection criteria (indeed it's included in the mortality risk analysis).

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:

Smith and Evans (2010) identify four methodologies to estimate the IEVSL: 1) meta-analyses of hedonic wage studies; 2) stated preference studies; 3) comparisons of VSL estimates at different points in time for a single country (Hammitt et al. 2003; Costa and Kahn 2004); 4) cross-country comparisons of VSL estimates (Hammitt et al. 2003). A fifth methodology involves the estimation of quantile regression such as in Kniesner et al. (2010) and Evans and Schaur (2010). The white paper emphasizes the first two methodologies. There are additional examples of 1. that should be considered including Mrozek and Taylor (2002), Viscusi and Aldy (2003), and Bowland and Beghin (2001). It would also seem reasonable to include estimates from Costa and Kahn (2004) given their focus on the U.S. in spite of the fact that they employ a methodology other than 1. and 2.

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?
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:

Robinson and Hammitt (2015), consistent with the selection criteria for the VSL white paper, exclude studies outside of the U.S. However, if the goal is to understand how the VSL varies with income, then expanding the analysis to include studies outside of the U.S. (which would increase the range of incomes over which we learn about the VSL), might be

reasonable. The selection criteria also focus on studies that provide estimates based on the U.S. population with studies of other subpopulations being less preferred (although Kniesner et al. is included and uses a sample of male heads of households). Given the goal of obtaining estimates of the IEVSL, it might be helpful to include other studies that use subpopulations but provide information sufficient to estimate the IEVSL.

There are also relevant studies outside of the environmental economics literature (e.g., Murphy and Topel, 2006).

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:

Between these two options, I prefer the second as I think the methodology should include results from stated and revealed preference studies.

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:

As noted in Robinson and Hammitt (2015), mortality and morbidity outcomes vary in many ways so “there is no reason to assume that the income elasticities are identical.” The recommendation to apply the IEVSL estimate to non-fatal health risks seems inconsistent with this.

Other minor comments/corrections:

1. Viscusi (2015) is cited on p. 57 but is not included in the list of references for the white paper.
2. Table 5—Smith et al. (2004) does use CFOI data but risk is measured by industry.
3. Table 5—The Health and Retirement Study is representative of older adults and when controls for selection into the labor force are included, the resulting VSL estimates should be representative of this segment of the population as well. Smith et al. (2004) includes such controls but Evans and Schaur (2010) does not.

Dr. Wayne Gray

To: Dr. Madhu Khanna, Chair EPA EEAC
From: Wayne Gray
Re: Preliminary Written Comments on Charge Questions
Date: March 3, 2016

I generally found EPA's White Paper "Valuing mortality risk for environmental policy: a meta-analytic approach" and Technical Memorandum "Income elasticity of VSL" to be a reasonable response to the earlier (2011) recommendations from the SAB-EEAC. Rather than repeating this general point for all of the charge questions, I'm only providing written responses in cases where I had more to say.

1b. Construct of the risk variable in hedonic wage studies.

I would have expected that studies using more detailed industry+occupation risk measures would find larger VSL estimates than those using industry-only risk measures, due to random measurement error in risk assignment. However, Viscusi (2004) finds that VSL estimates based on industry-only risk are roughly twice as large as those based on industry+occupation, suggesting some other mechanism than simple measurement error. Not knowing the nature of this mechanism makes it difficult to evaluate other approaches to generating within-industry variation, such as by gender and age. If it were simply due to measurement error, the gains from within-industry variation would be tied to the differences in risk across the groups. I would expect that occupation provides much more within-industry variation than either gender or age; it could be useful to explore whether studies using risk measures other than industry+occupation show systematically different results.

10. Inclusion of an influence analysis.

Yes, I think it's important to include an influence analysis. While the methodology used to pool the results from various studies seems reasonable, the complexity of some of the weighting procedures make it helpful to see whether the results from any given study greatly affect the overall (pooled) estimate.

11. Protocol for revisions over time.

I'm concerned that the reliance on published, peer-reviewed studies may be at odds with the desire to regularly update the VSL estimates to reflect changes in income (and possibly changes in preferences or labor market structure) over time. The publication process emphasizes novelty and provides little incentive to repeat earlier studies. To take one example, Viscusi (2004) examined the impact of using industry+occupation rather than industry-only measures of fatality risk, using FCOI data from 1992-1997. It would be valuable for the present purpose to know how those estimates would change if applied to 2002-2007 (or 2012-2017) data, but it's not clear that journal editors would consider those results especially newsworthy. That could make it

problematic to systematically exclude older studies, if there are no comparable newer studies to replace them.

12. Expanding the range of results considered for cancer risk.

It seems problematic to expand the range of studies considered for this particular item beyond those used to construct the primary VSL estimates. Certainly studies from other countries could reflect very different attitudes towards risk and the influences of different institutions (e.g. labor market, health care, and insurance) and should be excluded. On the other hand, risk-risk studies would seem well-suited to this sort of comparison, and could be considered. I'm somewhat concerned that the elevated responses to cancer risk in stated preference studies might be due to perceptions of that disease that would also apply to other fatality risks from pollution exposure (as compared to risk of sudden accidental death), so that differentially weighting cancer risk might be inappropriate.

14. Dealing with near-zero income elasticity estimates.

Given that we're willing to exclude estimates of income elasticities that are zero or negative, it seems odd to include values that are very near zero. Since the balanced approach focuses on the central tendency among a large set of studies, the treatment of a few studies with outlying results relative to our prior beliefs (either too high or too low) is less crucial. If a large share of studies provides such low estimates, it raises some concern about the general approach being taken to estimate income elasticities – since it seems implausible that avoiding fatality risk is not at least a normal good.

16. Choose between two alternatives for the income elasticity of VSL.

I'd prefer using the central tendency from Robinson and Hammitt combining both hedonic wage and stated preference results, rather than relying only on the hedonic wage studies examined by Viscusi. It seems more in keeping with the approach used to generate the primary VSL estimates. The results of the two approaches (0.7 vs 1.1) show that hedonic wage and stated preference studies are giving quite different answers to the question, which should be recognized as indicating some uncertainty in the IEVSL.

17. Estimating income elasticity for non-fatal health effects.

It seems reasonable to use the same income elasticity for both fatalities and non-fatal health effects. It avoids problems with relying on limited data to distinguish between the two effects, and also avoids the complication that the relative importance of the two risk reduction benefits for a given policy would diverge over time simply due to trends in incomes.

Dr. F. Reed Johnson

Preliminary comments on charge questions

Reed Johnson

1a. Evidence of validity for stated preference studies

EPA's White Paper uses a minimum standard of passing a weak-scope test for validity—which simply requires that empirical estimates of risk aversion have the correct sign and be statistically significant. It is worth noting that recent FDA draft guidance on submitting stated-preference data for regulatory benefit-risk assessments requires that data satisfy scientific standards for evidence.³ Among other things, these standards include:

- Well-informed respondents
 - Attributes defined relevant to clinical endpoints
 - Common understanding of benefits and risks
- Controlled experiment
 - Utility-theoretic measurement
 - Randomization of experimental stimuli
- Representative and adequately powered sample
 - Condition severity
 - Treatment and outcome experience
- Statistical analysis of effect sizes
 - Data-quality evaluation
 - Correct statistical modeling
 - Formal hypothesis testing

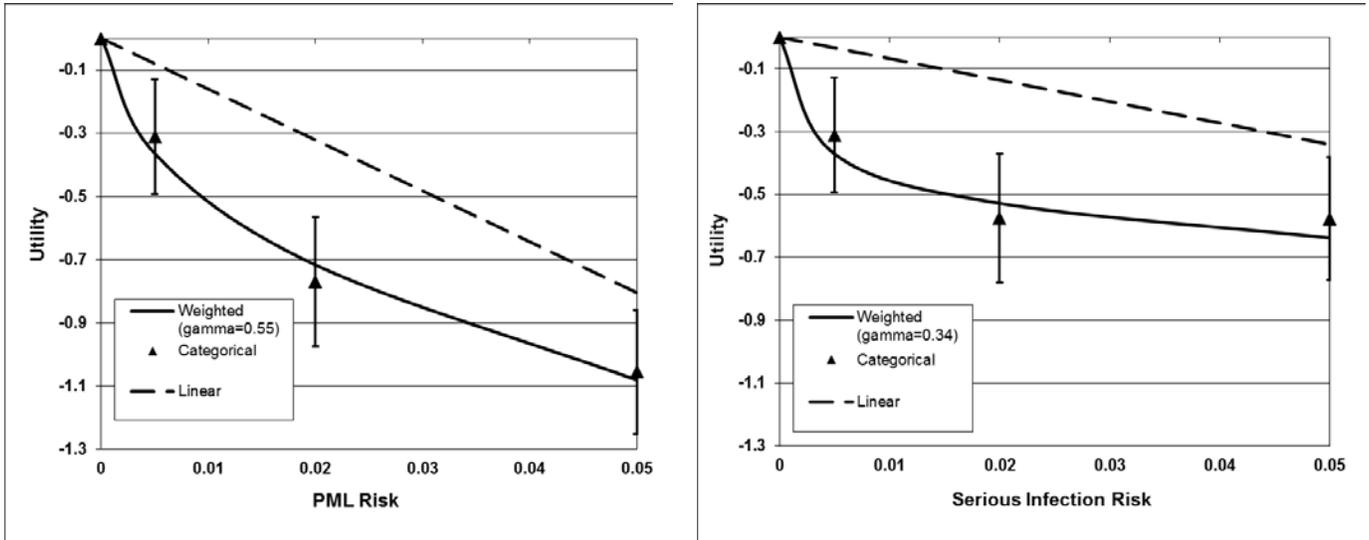
The theoretical basis valid VSL estimates generally relies on results derived from expected-utility theory. However, the expected-utility hypothesis has been widely rejected in empirical testing. Van Houtven, Johnson, et al. derive maximum acceptable risk welfare measures from a general utility-theoretic framework and reject the expected-utility hypothesis in favor of prospect-theory and rank-dependent utility specifications in an empirical application.⁴

This study also finds significant differences in risk aversion for different mortality risks. The following charts compare preferences for 10-year risks of death from treatment side

³ U.S. Department of Health and Human Services Food and Drug Administration Center for Devices and Radiological Health and Center for Biologics Evaluation and Research. Patient Preference Information – Submission, Review in PMAs, HDE Applications, and *De Novo* Requests, and Inclusion in Device Labeling: Draft Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. May 18, 2015.

⁴ Van Houtven G, Johnson FR, Kilambi V, Hauber AB. Eliciting benefit-risk preferences and probability-weighted utility using choice-format conjoint analysis. *Med Decis Making*. 2011;31(3):469-80

effects: serious infection and PML, a viral disease that is a side effect of some biologic medications.



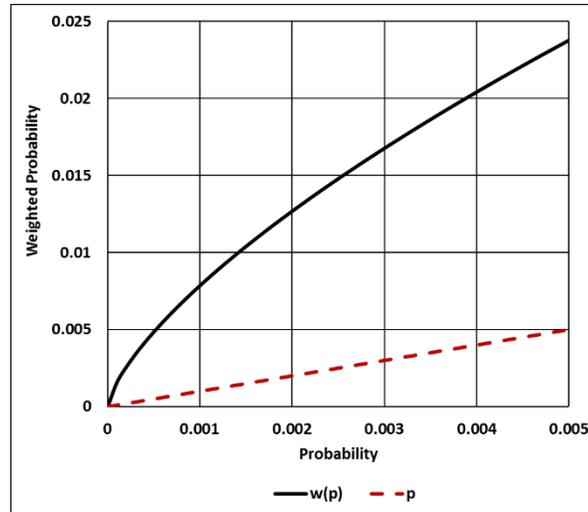
Source: Van Houtven, et al. 2011

The categorical models clearly are consistent with continuous models with weighted probabilities of the prospect-theory form proposed by Tversky and Kahneman for financial risks.⁵

$$\pi(p) = \frac{p^\gamma}{[p^\gamma + (1-p)^\gamma]^{\frac{1}{\gamma}}}$$

Weighted-probability functions that have been empirically validated typically have an inverse-S shape, with probabilities close to zero over-weighted and probabilities close to one underweighted. For small environmental-health risks, respondents to stated-preference surveys thus actually evaluate a different level of risk than that shown in the survey, as illustrated in this figure for $\gamma=0.7$.

⁵ Tversky A, Kahneman D. Advances in Prospect Theory: Cumulative Representation of Uncertainty. Journal of Risk and Uncertainty 1992;5:297–323.



This distortion does not affect WTP, but it does affect VSL when WTP is divided by the unweighted probability which biases VSL upward.

Thus concerns about validity are rather more complicated than traditionally discussed in nonmarket valuation of environmental health risks.

1b. Construct of the risk variable in hedonic wage studies

It was the previous panel's judgement that industry-level data were too heterogeneous to provide meaningful data for income-risk tradeoff estimates.

1c. Estimates for immediate risk reductions

The risk of immediate death is of limited relevance for environmental policy—corresponding presumably only to a massive cardiac infarction or stroke. It has been used in stated-preference studies because it is simple to explain to respondents. Matching revealed-preference data based on accidental deaths is logical, but simply compounds the weak policy relevance of the SP data.

It is worth asking what the likely bias is for the wider range of fatal and nonfatal health damages associated with pollution exposures. On the one hand, most real deaths involve some period of discomfort and disability, while pollution-related deaths may occur only with some lag or the result of cumulative exposure over time.

2. Comment on relevant empirical studies

I am not aware of any additional studies. The lack of significant growth in policy-relevant research is a major concern.

3. Comment on the population-weighting approach.

4. Methods used in the White Paper to estimate standard errors.

5. Application of meta-analytic methods to derive generally applicable VSL estimates.

The previous EEAC specifically recommended that “EPA work toward developing a set of estimates of VRR corresponding to policy-relevant contexts defined by the type or characteristics of the risk (e.g., associated morbidity, latency) and of the affected population (e.g., age, health, income). Economic theory and empirical evidence suggest that WTP can vary with these characteristics and that a single value of mortality risk reduction is not appropriate for all contexts.”

Six years later, contrary to this recommendation, EPA has instead chosen to implement a meta-analysis to obtain a single universal VSL value. What is the reason for rejecting the previous recommendations? I am concerned that this EEAC is being asked to endorse an approach that the previous panel explicitly rejected. The 2010 EEAC suggested 4 possible approaches: “(1) using only primary estimates obtained for the specific [policy] context; (2) developing adjustment factors to transfer estimates from other contexts; (3) developing meta-regression equations; and (4) structural benefit-transfer methods to characterize appropriate values across multiple contexts.” It is obvious from the report’s repeated concern about identifying policy-relevant values that the meta-regression alternative was not intended to suggest support for using this approach to identify a single VSL value.

To the extent the panel agrees to the terms of the Charge, my expertise in meta-analytic statistics is dated and I will defer to more knowledgeable members of the panel. However, why did EPA choose an error-components analysis requiring imputation of unobservable parameters instead of a random-effects model that assumes (albeit incorrectly) that all observations are sampled from the same underlying distribution? In that case, the contribution to the likelihood function of a sample mean and variance is:

$$L_i(b_i, s_i^2 | \beta^*, \sigma^{*2}) = \left[\frac{1}{\sigma^*} \exp\left(-\frac{n_i(b_i - \beta^*)}{2\sigma^{*2}}\right) \right] \cdot \frac{1}{\sigma_i^{*n_i-1}} s_i^{n_i-3} \cdot \exp\left(-\frac{n_i s_i^2}{2\sigma^{*2}}\right)$$

where n_i , b_i , and s_i^2 are the number of observations, sample mean, and sample variance, respectively and β^* and σ^{*2} are the population mean and variance.

6. Accounting for potential correlation of results that rely on the same underlying data.

7. Optimal weights that account for sampling and non-sampling errors.

8. Non-parametric and parametric approaches.

9. Use of estimates from the non-parametric model, balanced, mean of study mean and parametric, balanced models.

10. Influence analysis.

11. Statistical criteria for standardized protocol for future updates.

In the 6 years since the last VSL White Paper, EPA has identified only 7 acceptable studies that were not available in 2010: 4 SP studies (3 using the same data) and 3 HW studies. Furthermore, the most recent data used in these studies were collected in 2009. There does not appear to be much urgency to develop a standardized protocol for future updates if increases in the literature base is about one study per year.

12. Selection criteria for identifying studies for estimating a cancer differential.

The previous EEAC report concluded that “research suggests that people are willing to pay more for mortality risk reductions that involve cancer than for risk reductions from accidental injury and proposes a placeholder value that could be used for this cancer differential while the Agency pursues long-term research to differentially value other types of risks.”

Six years later, EPA still finds the available VSL literature inadequate for informing policy on differential health effects. The Charge cites the previous EEAC recommendation that “...EPA work toward developing a set of estimates...for policy-relevant cases characterized by risk...”. Furthermore, the previous EEAC stated that the Agency should “encourage research to obtain revealed and stated preference estimates for the types of risk and types of affected populations that are most relevant to environmental policy contexts.” That obviously has not happened. The Agency has not provided support for research on policy-relevant estimates of the public’s value for reducing environmental health risks. Reliance on funding from other sources and the waning interest in nonmarket valuation relative to other research topics in environmental economics is unlikely to result in a significantly richer research literature to discuss when the next EEAC convenes 6 years from now.

13. Summary of the income elasticity of VSL and non-fatal health effects.

I was responsible for the income-elasticity recommendations in the previous EEAC report. My view of the problem has not changed. The report says: “The decision on how to adjust VRR for income growth over time is not independent of what approach is adopted to support the VRR value or range of values. The committee recommends selecting studies that are matched as closely as possible to the policy context in question. When closely matched studies are not available, then the committee

recommends developing a benefit-transfer function. This recommendation implies that the simple, two-period, constant relative risk aversion model is an inappropriate guide to deriving income elasticity. Specifically, this model does not allow VRR covariates to influence income elasticity, assumes homogeneity in time preferences, and requires mortality risks to be homogenous and exogenous.”

14. Income elasticity estimates that are quite low.

From the previous EEAC report:

“The literature on VSL income elasticity has employed several approaches, including cross-section analysis of within-sample variation in CV data, meta-analysis of hedonic-wage studies, longitudinal analysis of hedonic-wage data for a particular population, and quantile analysis of hedonic-wage data. Unfortunately, stated-preference estimates that are most closely matched to the policy context may lack the necessary information to derive a utility-theoretic elasticity estimate. ... Elasticity estimates generally vary with age and income, with the inverse relationship with income being the stronger effect. ... Thus smaller elasticity values may be biased against lower-income groups.”

15. Selection criteria.

From the previous report:

“As indicated in the committee response to the other charge questions, the committee does not believe EPA’s focus on a single value or range of values is appropriate. Both VRR and income-elasticity values should be context specific. Where policy-relevant estimates are lacking, EPA should develop income-elasticity transfer functions that make any necessary assumptions for interpolating or extrapolating the available evidence transparent and that facilitate appropriate sensitivity and uncertainty analysis.”

16. Alternatives for arriving at IEVSL estimate.

From the previous report:

Consistent with its recommendations on VRR, the committee recommends that EPA develop an elasticity transfer function that accounts for changes in age and income distributions over time. Policy impacts that affect particular regions or populations should account for differences in the age and income distributions of the affected populations relative to the national distribution.

17. Using the IEVSL for the value of non-fatal health risks.

Dr. Matthew Kotchen

DATE: March 1, 2016

TO: EPA Science Advisory Board (SAB) Environmental Economics Advisory Committee

FROM: Matthew Kotchen, Yale University (committee member)

RE: Preliminary written comments for discussion with fellow committee members in preparation for the March 7-8 meeting on the EPA's proposed methodology for updating its mortality risk valuation estimates for policy analysis

Question 1. Selection criteria for study inclusion.

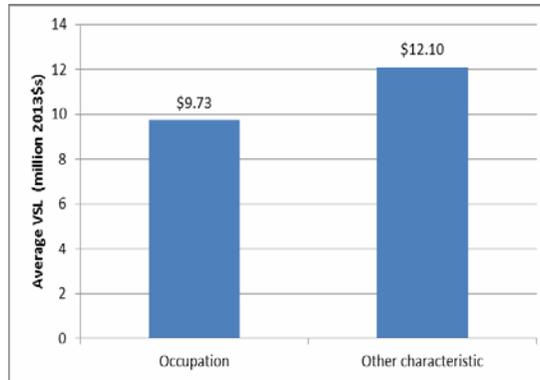
The EPA appears to have responded appropriately to the recommended selection criteria for study inclusion.

Question 1a. Validity of stated preference studies

I agree with the previous committee recommendation that any included stated preference study “should provide evidence that it yields valid estimates.” I believe the EPA has interpreted this guidance in a way that is appropriate and scientifically sound. Appendix B of the White Paper documents the procedures in a clear and transparent way. That the stated preference results tend to be lower than the revealed preference results builds confidence that hypothetical bias is not a concern for the present analysis.

Question 1b. Construct of the risk variable in hedonic wage studies.

In principle, I support the previous committee recommendation to “eliminate any study that relies on risk measures constructed at the industry level only (not by occupation within an industry).” This will minimize the potential for unobservable industry-level effects to bias the estimates of risk on wages. The EPA's interpretation of the recommendation to allow studies where the risk measure differs by at least one other characteristic—that is, variables such as gender or age, in place of occupation— appears reasonable given the small number of studies that consider occupation. This increases the number of studies from 4 to 9 (though text in the report references 8). To see the implications of broadening the criteria, the following figure shows the VSL estimates from studies that use data differentiated by occupation or other criteria (2013\$s from Table 6, average across studies after averaging within studies and using information in Appendix B). It appears that broadening the assumption tends to increase the VSL estimate. This information may be important when considering EPA's approach.



Question 1c. Estimates of the average value of the marginal WTP for immediate risk reduction.

I believe the EPA has used appropriate methods for estimating immediate risk reductions.

Question 2. Adequacy of the empirical studies.

To the best of my knowledge, the EPA has adequately captured the relevant stated preference and hedonic wage literatures. I have no additional studies to recommend.

Question 3. Appropriateness of the population weighting approach.

The population weighting approach appears appropriate and scientifically sound.

Question 4. Methods used to estimate standard errors.

The approach to estimating standard errors appears appropriate and scientifically sound.

Question 5. Meta-analytic methods.

The White Paper appears to represent an appropriate and scientifically sound method for deriving estimates of the VSL for purposes of policy analysis based on the best available research. There has been a careful screening of studies to include and the statistical methods are clear, with assumptions explicit, and the results are reasonable.

Question 6. Grouping samples for the analysis.

Yes, the grouping of studies is appropriate to account for non-independence of different estimates used in the meta-analysis.

Question 7. Optimal weights that account for sampling and non-sampling errors.

I believe the weighting approach is appropriate.

Question 8. Non-parametric and parametric approaches.

These are appropriate methods and provide a reasonable range of estimates. I do not have any alternatives to suggest.

Question 9. Conclusions from the analysis.

I appreciate the different options presented and discussed in the paper. I support the EPA's judgment about the preferred estimates, though I am open to further discussion with fellow committee members.

Question 10. Influence analysis.

I think the influence analysis is well-done and contributes to the report. As a general rule, it is useful to see if any one study is having an outsized influence on the results. None of these strike me as sufficiently large to suggest omission. The results also show how the balanced estimates are generally less swayed by an individual study, lending support to the EPA's preferred estimates.

Question 11. Threshold for commencing new updates.

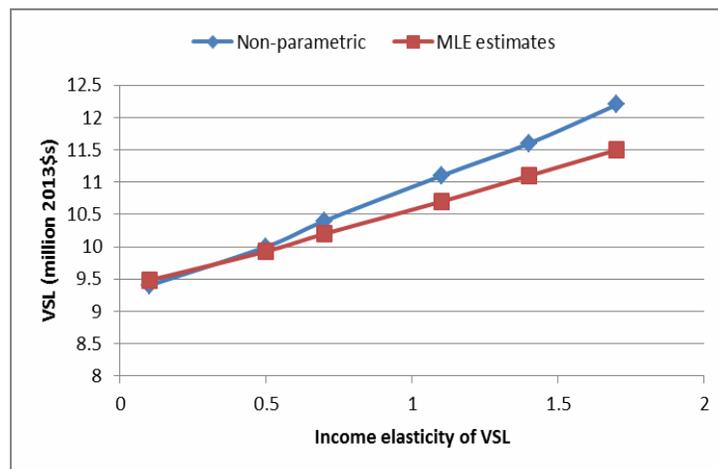
A schedule of updates every 5 years seems appropriate. Repeating the same methods for the inclusion of new studies is appropriate. Establishing some duration for dropping studies might also be appropriate, perhaps 20 years. But perhaps the influence analysis can be useful in this regard as well. As long as an older study still meets the criteria of inclusion and is not an outlier, inclusion might still be justified.

Question 12. Comparison of values for cancer and non-cancer risk reductions.

At the present time, there does not appear to be enough information to identify a cancer differential that should be used for differential policy analysis. Perhaps the question can be revisited during the periodic reviews that are the outcome of the previous question.

Question 13. Summary of the income elasticity and the literature on it for non-fatal health effects.

I believe this is a reasonable summary of the literature and provides a solid basis for making a determination about the income elasticity of the VSL. I am not aware of any other studies that should be included. Regarding the non-fatal health effects, I wonder if there is anything in the stochastic/dynamic literature on savings and labor that would be informative here. I hope to look into this a bit more. To get a sense for the importance of the parameter, I use data from Appendix 3 to produce the following plots on the relationship between the IEVSL and VSL estimate using the EPA's balanced approach:



Question 14. Addressing very low, non-zero mean reported income elasticity results.

I support the approach used in the Technical Memorandum.

Question 15. Study selection criteria in income elasticity report.

The study selection criteria seems reasonable.

Question 16. Alternatives for arriving at an income elasticity estimate.

I support using Option 1 (equally weighted mean results from the HW and SP literatures). This is in parallel with the estimates from the VSL approach, and it does not place too much emphasis on one researcher's study. It also has the advantage of being straightforward. I do not support using the other approach for sensitivity analysis, as it would be preferable to use the bounds produced from option 1 directly.

Question 17. Using income elasticity of mortality to estimate elasticity for morbidity.

This is a difficult question because I think there are important theoretical issues on the relationship between these parameters, and I am not sure they have been worked out. I suspect using the IEVSL statistic for non-fatal health risks is only accurate for a very limited special case. Nevertheless, it is unclear what should be done because a value of zero is more surely incorrect. Hence I support using the IEVSL until more research is done in this area, but it should be a priority area for EPA research and/or support. This recommendation also reflects the fact that there is little difference in the other approaches outlined in the Robinson and Hammitt (2015) report. Although a committee discussion on this topic will be important.

Dr. Matthew Neidell

Comments for charge questions for March 7 & 8 SAB meeting
Matthew Neidell

1a. The methods used to assess validity generally seem like reasonable tests, but it is hard to know why some papers pass the test and others don't. Do the ones that don't pass have a design flaw or something that makes them different from the ones that pass? Or are the ones that pass "lucky" to have passed?

1b. I take this to mean that it excludes studies with risk measured at the industry level only. However, some of the risk calculations need defense. Why does the risk vary by race within an industry (Viscusi, 2003)? Are certain groups more protected or is this a proxy for occupation within an industry?

1c. It seems appropriate to include these additional studies. Much of the mortality risk from environmental exposure is not immediate, making it natural to include studies without immediate risk. Discounting may be controversial, however; see comment 15 below.

2. It is not clear to me why RP studies that aren't necessarily hedonic wage studies are excluded. Two that come to mind:

Ashenfelter, Orley and Micahel Greenstone. "Using Mandated Speed Limits To Measure The Value Of A Statistical Life," *Journal of Political Economy*, 2004, v112(2,Part2), S226-S267

Davis, Lucas. "The Effect of Health Risk on Housing Values: Evidence from a Cancer Cluster." *American Economic Review*, 2004, 94(5), 1693-1704

Although Davis' study is limited in geographic coverage, I imagine weighting can be done to make it nationally representative, as with other studies.

Additionally, here is another HW study not included:

Deleire, Thomas, Shakeeb Khan, and Christopher Timmins. "Roy Model Sorting and the Value of a Statistical Life," *International Economic Review*. Vol.54, No.1 (2013):279-306.

3. The weighting approach seems appropriate and scientifically sound.

4. When standard errors of the VSL are not present, the White Paper typically uses the standard error of the risk coefficient. This approach seems generally appropriate.

5. The meta-analysis generally seems appropriate, but would benefit from citations to support the choice of various equations.

6. I am concerned that too much weight is being placed on multiple estimates from the same paper, especially given assumptions needed to estimate non-sampling variance. I would like to see, as a point of comparison, estimates using only “central” or “baseline” estimate from each paper.

7. I would like to see a reference for equation 4 given that this is central to this analysis. It seems appropriate, as do the steps to operationalize it, but I have not attempted to derive the properties of these estimators, some of which are deemed “consistent.”

8. The parametric and non-parametric approaches seem reasonable, though again I would like references to support it.

9. I am troubled by the conclusion that the “mean of group estimators” is the preferred estimator (this relates to comment 6 above). It seems essential to incorporate standard errors into the group estimator.

Another concern, which relates to comment 1 above, is why are the SP estimates lower than the HW? Since the SP studies reflect things like cancer risk as opposed to occupational risk, it would seem as though the estimates should be larger, if anything. Also, why is there so much variance in the SP estimates?

10. The influential outlier analysis seems appropriate, though a citation would be useful. Even if appropriate, not clear to me why it is needed.

11. For the frequency of updates, assess if new papers are available once a year. If they are, incorporate into existing model and re-estimate. For exclusion criteria, it seems as though we can add more papers by a) including RP studies outside of the labor market and b) including studies that are not representative but can be weighted to become representative.

12. Seems like a slippery slope to try to separate cancer. Why separate cancer only, and not other diseases, like COPD and cardiovascular disease? Both have a latency period with years of morbidity, much like cancer.

13. The approach generally seems acceptable. One main questions, though see specific comments below, is: why not use difference in income within a given time period (across people) to estimate the income elasticity? I’m not aware of any additional studies.

14. All of these near zero effects come from SP studies. This raises the concern that some of the validity tests imposed on SP studies may not be sufficient. As with above, it would be useful to see an analysis that completely excludes SP.

15. The methods seem clearly enumerated. The 10-year delay in mortality in the Viscusi, Huber and Bell add an element of discounting, which distinguishes it from other papers. This makes the income elasticity harder to obtain, especially if hyperbolic discounting exists. A justification for excluding it.

16. Rather than IE estimates specific to VSL, why not use general income elasticity of demand estimates from the literature more broadly? The value from additional wages comes through the ability to consume goods and services.

17. This approach seems oversimplified since the effect of mortality operates differently from the effect of morbidity (in a Grossman-type model). What about using the income elasticity of demand for health care instead? Or see comment 16.

Dr. James Opaluch

Charge Questions

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.

Validity is a multi-dimensional criterion and evidence for validity lies along a continuum. It would be desirable to have a set of validity criterion and metrics. Protocols should be established, and authors should be encouraged (or required) to adhere to protocols. Generally, more rigorous testing is associated with the validity tests along the continuum, and one must ask how selective the criteria should be to justify inclusion of a study. The tradeoff, of course, is the more demanding the criteria, the fewer studies would qualify, and thus be available for estimating value of risk reduction.

In general, EPA employs proper validity criteria described in Appendix B of the White Paper. The first validity criterion not explicitly discussed but is implicit in Appendix B is that the studies being adopted employ methods that are recognized to provide proper measures of risk preferences (hedonic wage and stated preference studies). This validity criterion requires that the studies provide measures the correct concept of the value of risk reduction. Related validity concepts include that correct “commodity” is specified (mortality vs. non-lethal injury vs. mortality plus morbidity) and that the instrument (e.g., questions in a stated preference survey) are “well-constructed” to measure risk preferences.

A second validity criterion implicit in Appendix B is that the coefficient on the risk measure not of the wrong sign—the study should not find a positive WTP for an *increase* in risk. A refinement of this validity concept is the “weak” scope test—as the study should find that a higher level of risk reduction should be valued higher than a lower level of risk reduction.

For study results to be logically consistent with the value of a statistical life concept, WTP should be linear in the level of risk reduction. This is because calculations in the VSL approach aggregates risk across individuals to calculate an estimated number of “lives lost”. For example, the VSL approach views the “commodity” as identical (1 statistical life) If a million people each face a risk of one in a million, or if 10 million people face a risk of 1 in 10 million. If WTP to reduce risk is nonlinear in the level of risk reduction, then the VSL concept does not apply. This is obviously a quite rigorous criterion that may be worth assessing, but I’m not sure I recommend studies be eliminated simply because they do not pass this

criterion.

Another form of validity properly considered in the White Paper and other papers is that other factors (income) be related to VRR in a way that is consistent with theory. One might expect people with higher income and those who drive more miles to have a higher WTP for safer roads.

At the same time, one should be caution of validity tests that are dependent upon specific results, since there is always a danger that “publication bias” might cause misleading results. For example, if risk coefficients that are not statically significant are automatically rejected, there is danger that an upward selection bias could result because studies that *overestimate* risk effects would be included in the analysis, but studies that understate risk preferences are excluded.

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.

As I read the SAB comment, it expressed the concern that risk measures constructed at the industry level cause measurement error because they pool together many different occupations that face substantially different risk (e.g., coal industry accountants vs. coal miners). I would argue that gender and age are not the relevant factors to be considered to solve this measurement error 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

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.

As I see it, there are at least two aspects to the question of cancer death in 10 years vs. immediate accidental death. First, there is the question of morbidity impacts, as discussed in the SAB comments. The cancer case includes both morbidity and mortality, while the accidental death involves mortality only. If one suffers instantaneous accidental death 10 years in the future, one has ten years of life at a status quo (no incident) level of quality followed by sudden death due to the incident. In contrast, death by cancer involves some rate of increasing pain and debilitation over the 10-year period, followed by death. Hence,

in the case of death by cancer, one suffers from morbidity and hence a deteriorating quality of life over the ten remaining years, while in the case of accidental death one faces the status quo for ten years, followed by instantaneous death. Simply discounting back to the present time is not the correct way to adjust for the differing quality of life that occurs in during the 10-year period. The SAB report correctly points out that it is essential to differentiate between the time lag to mortality in the cancer case, versus the time lag to onset of symptoms.

Another issue arises as to how to translate death 10 years from now compared to instantaneous death today. This is different than receiving a monetary payment ten years in the future, where discounting is appropriate, and where one could conceivably borrow against. But dying today (vs 10 years from now) means losing 10 years of life. You dying immediately means you get less years of life, not just a delay.

This aspect of the problem is better framed as years of life extended (or years lost), rather than saving (or losing) life. If an individual does not die from a particular risk, the person will eventually die from some other cause at some point. Avoiding this one risk means lowering the probably of death at some point in time, but all other risks remain. This is different than receiving a sure payment some particular time in the future when I can expect to live to that time with probably (close to) one. A twenty-year-old might willing to pay a considerable amount to reduce a risk faced 30 years in the future. In comparison, a 70-year-old may be willing to pay little or nothing to avoid a risk of death 30 years in the future. This means the value of risk reduction depends upon the kinds of risks faced, and the population that faces that risk. The value of reducing automobile deaths may be worth more than sources of mortality that primarily effects the elderly, since they disproportionately affect young people. Avoiding a randomly selected automobile death leads to saving more years of life than avoiding the death of a randomly selected nursing home occupant.

To make the analogy closer, compare dying instantaneously 10 years from now versus dying 20 years from today. This is more akin to receiving a 10-year stream of declining income payments starting today, versus receiving a 10-year stream of declining income payments starting 10 years from now. This is quite different than the problem of comparing a fixed sum (\$1,000) ten years from now versus delaying that same fixed sum for 20 years. It is not merely that the payment delayed, but rather the time stream of payments is different. In the same sense, dying 10 years from now provides a time stream of 10 years of life, not merely a delay in a fixed outcome.

If the goal is to measure mortality risk only, then the least ambiguous measure concerns immediate death. If examines death by cancer ten years in the future, the individual might undergo pain and suffering for part of that time, dying in 10 years. Simply discounting back to the present time does not correct for the pain and suffering.

At the same time the concept of saving a life is not the correct way to frame the problem. The fact is, the morality rate is 100%. It is not a question of whether we die or not, but *when* we die. If someone told me that undertaking an activity will cause instantaneous death in 50 years with 100% certainty, it might not place any value whatsoever on that particular risk because it is unlikely to have any effect on my lifespan. I will most certainly die from some other cause before that time.

Note that this is different than the issue applied at the population level. In this case, a risk that occurs 10 years in the future represents not delay in the death of an individual, but a delay in risk faced by an overlapping, but not identical set of individuals. This means we need to deal with the delay in death caused by an incident today differently than a delayed risk that is faced 10 years from now. The proper way of framing this problem is one of years-of-life extended, not loss of life. If an incident occurs at a point in time which results in delayed mortality, the individual loses fewer years of life, and simple discounting is not the proper way to correct for the delay. In contrast, if there is a risk of an incident (e.g., accidental death) 10 years from now that causes instaneous death at that time, the pool of

individuals facing the risk is different. Some individuals facing a risk 10 year from now are not yet born, while some individuals facing the risk today will be dead 10 years from now. Ignoring for the moment any differences in the characteristics of the population at risk today vs 10 years from now (e.g., life expectancy, age structure of the population, real income, etc.) there is an identical number of years of life lost, and therefore discounting is proper.

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.

Correcting sampling errors for non-representative subpopulations is a challenging problematic, since it is not feasible to correct for all non-representative characteristics of the sample. Surveys commonly find that better educated and wealthier individuals are more likely to respond to surveys than less well educated, poorer individuals. These types of sampling errors can be easily corrected by using weighting subsamples to be representative of the general population. But there are other factors in selection bias that are not so simple to correct. For example, people more interested in the topic of interest may be more likely to respond than individuals who are less interested. But it is not straightforward to compare the level of interest of the sample vs population in order to develop the weights necessary to correct for sampling error. There may be many other factors that are even more challenging, such as personality traits, etc. that are even more challenging to identify, measure and develop weights. But no analysis can be perfect.

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.

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.

The two extreme cases would be to take a simple average of all estimates from all studies, versus averaging the average estimates of the studies (mean of means). Ignoring for the moment other differences across studies, the former approach is appropriate if the estimates within a study are statistically independent, while the latter approach is appropriate if the estimates within a study are perfectly correlated. The two

estimates provide a reasonable sensitivity analysis for the importance of considering within-study correlation. My personal opinion is the mean-of-means, based on perfect correlation, is probably the more appropriate of the two. If sufficient data were available, the variance component approach described in the white paper is the preferred approach. The mean-of-means is probably the best pragmatic approach given the shortcomings of the data (e.g., some studies have only one estimate, so cannot be used to estimate the variance component). It might also be that a fixed effect model is more appropriate than a random effects, variance component approach. This is especially relevant since there is no reason to expect that the studies are drawn from the same sort of distribution (e.g., have constant within-study variances). Robustness of fixed effects vs. random effects models becomes an important consideration (see discussion to the following charge question)

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.

The variance components approach discussed in the White Paper is an interesting framework. However, given data limitations is probably not practical. In addition, as discussed above, the random effects approach assumes that the studies are drawn from the same distribution (e.g., same within-study variances). This assumption is not likely to be true given differences in study methodologies, and one would like to know the robustness of random effects models when the underlying distributions are not the same.

Some literature suggests that random effects models may be robust to many sorts of misspecifications, at least in cases of large data sets (see references below). My understanding is that fixed effects models tend to be preferable with small datasets, as is the case in the White paper, while random effects models are more appropriate for larger datasets. Having said that, it is definitely problematic to use a study-based fixed effect when there is only one estimate from a particular study.

Charles E. McCulloch and John M. Neuhaus, 2011. "Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter", *Statistical Science* Vol. 26, No. 3, 388–402.

Vignoles, Anna and Crawford, Claire and Steele, Fiona and Clarke, Paul (2010) The Choice Between Fixed and Random Effects Models: Some Considerations For Educational Research. In: British Education Research Association Conference Special session on School effects on pupil outcomes, quantitative methods and applications, 1-4 September 2010, University of Warwick. (Unpublished))

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.

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.

It is interesting and important to provide a range of estimates or sensitivity analysis based on different approaches when no single approach is “correct”.

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.

I believe the EPA use of influence analysis provides valuable perspective on the results. Some form of Influence analysis is important for meta-analysis in cases where there are few studies to consider, and therefore one or two individual studies might have a substantial influence on the estimates. Most significantly for influence analysis is that the influence not skew the results in a single direction. For example, if there are two studies with +10% and -10% the two studies are more or less balanced. Looking at the mean of group means, the two most influential studies are Corso Hammitt and Graham (-13.8%) and Chestnut, Rowe, and Breffle (-11.1). Taken together, these studies pretty much balance each other. In contrast, for the maximum likelihood SP, the Corso, Hammitt and Graham (-22.8) is well over 2 times the 2nd most influential study, which fortunately is of the opposite sign. Rather than dropping CHG altogether, one might use a robust estimation technic that limits the influence of this observation—one possibility is to adjust the weight on this study downward until it just balances the Alberini et al study, or to downweight all studies that are identified as relatively influential (perhaps studies that fall above the +/- 10% range?). This type of approach of downweighting highly influential observations has a long history.

In general, I would say that individual studies are not overly influential considering the small sample size, with the one exception being the Corso, Hammitt and Graham SP results when using maximum likelihood. The EPA approach provides a good perspective to understanding highly influential studies.

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.

It is essential that the profession develops a set of protocols for empirical analysis in order to support credible meta-analysis. One issue is developing protocols for deciding when to update the EPA meta-analysis, and which studies to add and/or remove. But EPA should also think about developing standardized protocols working with other federal agencies that can ensure economic studies that are supported by federal funds are of higher value to the nation.

Clearly, meta-analysis is essential if economic studies are to contribute broadly to public policy decisions. It is simply not feasible to fund *de novo* studies for each and every policy question to which economic analysis can provide important input. Hence, some sort of meta-analysis of existing studies is essential to use economic studies to guide policy in an efficient way. But if economics studies are used to guide

policy, we must take the full process of data collection, analysis and meta-analysis very seriously. This means providing sufficient information to assess results from individual studies. This includes, for example, a requirement that full copies of surveys be made available to readers, most likely through supplementary materials.

Similarly, we must develop careful reporting protocols so readers and other researchers can better utilize the results. This certainly means providing such statistics as standard errors of significant metrics (e.g., estimated VSLs). Developing protocols requires a thoughtful analysis in itself, perhaps to be carried out by an SAB committee. Fortunately, we can draw from the rich experience of colleagues in other disciplines with far more experience, especially in medical research.

Federal agencies including EPA, NOAA, NSF and others can play an important role in ensuring this happens. First, agency-funded studies should be required to conform with specific reporting standards. Some important first steps have been made in this direction through requirements like Data Management Plans, but there should also be specific reporting protocols for studies that are supported by federal dollars. At a minimum, studies should report such things as standard errors of estimates, but this is

More generally, studies should be required to provide documented data that are made available to the agencies, subject of course to careful confidentiality standards. While not understating the challenges, access to study data would allow

Data whose collection was funded by Federal dollars should be considered public property to be made publically accessible, with obvious exceptions for data that is classified, sensitive and/or confidential. But public availability of data should be the rule for data collected with federal funds, and restrictions on data should be the exception. National Institutes of Health has adopted the policy:

“In NIH's view, all data should be considered for data sharing. Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data. To facilitate data sharing, investigators submitting a research application requesting \$500,000 or more of direct costs in any single year to NIH on or after October 1, 2003 are expected to include a plan for sharing final research data for research purposes, or state why data sharing is not possible.”

Development of data management plans are an important first step. But data storage and documentation also needs to be more standardized if it is to be used for meta-analyses.

We have much to learn from our colleagues in other disciplines, and such as the Geographic Information Systems community, which has standard metadata protocols such as the Content Standard for Digital Geospatial Metadata (CSDGM) (<https://www.fgdc.gov/metadata/geospatial-metadata-standards#csdgm>). Federal agencies should consider the possibility of designing research contracts that specify that the deliverables including rigorously documented datasets based on carefully designed metadata protocols. Clearly this will be a significant challenge, but with datasets in hand, new policy-relevant analyses can combine data from previous studies in a much more thoughtful and rigorous manner.

Similarly, peer reviewed journals should also consider requiring data sharing. For example, Nature including the following policies on data, materials and methods underlying papers:

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature journal is that authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must also be disclosed in the submitted manuscript.

After publication, readers who encounter refusal by the authors to comply with these policies should contact the chief editor of the journal. In cases where editors are unable to

resolve a complaint, the journal may refer the matter to the authors' funding institution and/or publish a formal statement of correction, attached online to the publication, stating that readers have been unable to obtain necessary materials to replicate the findings. (<http://www.nature.com/authors/policies/availability.html>)

Nature also includes specific requirements for authors to make data available (<http://www.nature.com/authors/policies/availability.html#data>).

More broadly, the economics profession needs a culture change, that includes careful data collection and information reporting. This must be led by the Journals requiring that submitted papers include adequate information to validate research results, and for agencies and other to put research results to better use. In biology and other fields of natural science, researchers can publish data that they collect. This develops a culture of careful data collection and provides rewards to those who collect high quality data. In comparison, in economics the primary professional rewards for academicians is for theoretical and methodological work. While there are clearly exceptions, empirical work published in refereed journals is primary for illustrating methodological work, rather than to answer empirical questions in an of themselves. The quality of data collection and reliability of empirical work could be greatly enhanced by a culture change that rewards careful measurement, data collection and empirical analysis.

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?

Carefully selected studies from the gray literature could be very beneficial. Indeed, there is no a prior reason to believe that gray literature estimates to be systematically of lower quality than that peer reviewed literature. Carefully peer reviewed gray literature studies should be included to extend studies in peer reviewed journals.

Similarly, high quality studies of risk-risk tradeoffs could be very useful for improving our understanding of differences between VSL estimates associated with different causes, and could provide an important role in validation. For example, differences in VSLs for different causes of death could be compared to stated preferences for different causes of death to provide a test of convergent validity.

Above I discussed the issue comparing of cancer death involving a combination of morbidity and mortality. One would expect a positive cancer differential for two death occurring at the same time t years in the future, where one death occurs instantaneously at year t , while the other involves suffering from cancer followed by death at year t .

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.

Income elasticities of VSL could be used for at least two different purposes: looking cross sectionally across affected populations with different incomes, and looking temporally at how VSL should be adjusted across time as income increases. The former will likely be a problem for ethical and political reasons. The latter might better be addressed simply by looking at how VSL changes across studies over time, or one might replicated a standardized VSL study in different time periods, similar to what is done for other studies, such as the National Survey of Fishing, Hunting and Wildlife Associated Recreation. If one had a time series of VSL using a single standard methodology, one could regress estimates across time with an income and time variable, and identify the extent to which VSL increases over time due to an income effect, versus some other time varying process. A simple time trend could be used in the absence of sufficient data to identify separate income and time effects.

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?

While I agree one would expect a positive income elasticity, there is a danger of eliminating low (zero) estimates, while not eliminating high estimated studies, as this decision rule tends to bias results. Similarly, eliminating “anomalously” low income elasticity estimates of .12 and .08.

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).

I fear that inclusion of the VHB article is a case where we include observations we like, and exclude those we do not like. I wonder whether it is more honest to simply assume income elasticity is one.

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?

In theory, it is problematic to drop estimates where income elasticity are low from the data set. If we think of each estimate as an observation of a random variable drawn from a statistical distribution, a selection bias results if we drop observations when the value is judged as “too low”, and don’t similarly drop observations when the value is high. In practice, one might argue that these observations fail a validity test, but this seems to me more a case where we drop observations we don’t “like”, and keep those that we do. This results in a form of selection bias. An influence analysis approach would argue for a more symmetric criterion, where high and low values are drop.

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.

I don't see any theoretically justifiable argument that the income elasticity for mortality risk should be the same as the income elasticity of mobility risk any more than the income elasticity for apples should be the same as the income elasticity for oranges or beef. Commodities are somehow related, but I am not aware of any reason that the income elasticities should be the same (nor different).

IEC Income Elasticity Report. I don't follow the argument from bottom of page 6 through the top of page 7. Is this simply saying it is an estimate of the vertical shift in the demand function, rather than the horizontal shift in demand? That is, the shift in the inverse demand function (p as a function of q and y)?

If an individual is willing to pay more for any given commodity, then when faced with a (continuous) set of options at different prices, a richer person would buy a larger "quantity" of the good. Richer people would pay more for a given level of risk reduction, and they would purchase more risk reduction when faced with a give set of options.

Dr. Andrew Plantinga

Charge question #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.

Preliminary Comments by Andrew Plantinga and Sylvia Brandt

Assessment of methods

The question about whether the methods “span a reasonable range of appropriate, scientifically sound, and defensible approaches” could have been answered, in part, by the authors of the white paper. There are no citations given for the estimators in 1-5 on pages 22-23. While the methods are clearly presented, they could be better justified in terms of the particular application—finding the central tendency of VSL estimates from studies that in most cases report multiple estimates.

Equation (1) is the object of interest and all five approaches represent alternative schemes for defining the weights w_{ij} . Estimator 1 and 2 are straightforward approaches but do not make use of information in the studies on sample size or sampling error variance. Estimator 1 (simple mean) is potentially problematic because it gives the most weight to papers with the most estimates. It is not clear why there is so much variation among studies in the number of reported estimates. Are the papers with a lot of estimates reporting robustness checks, demonstrating differences in estimates resulting from specification changes, or something else? Either way, it is not clear that multiple estimates from a single paper should be treated as if they come from separate papers. Estimator 2 (mean of group means) has the potential advantage that it gives the same weight to each paper.

The sample size weighted mean (estimator 3) uses information on sample size and appears to be a standard approach (it is described in the text on meta-analysis by Hunter and Schmidt (2004)). Likewise, the sampling error variance weighted mean uses information on the sampling error variance. It is described in another text on meta-analysis (Hedges and Olkin 1985) and implemented in a recent meta-analysis by Hsiang et al. (2013).

Estimator 5 involves directly estimating the group and observation-level non-sampling errors. The estimate of the individual non-sampling variance requires that the total variance of each estimate, $\text{var}(y_{ij})$, be estimated from within-group variation (equation 5). There are only four studies that provide more than four observations and so these estimates are likely to be very imprecise. The authors take a weighted average of these estimates (equation 8) to obtain a single

individual-level non-sampling variance estimate. Other than practicality, it is hard to see a justification for this assumption. Our assessment is that this approach makes sense if there are a large number of estimates from each study.

The parametric estimation includes study characteristics: year data collected, SP or HW, median. It is not clear what the median dummy captures, although our guess is that it signifies that a median estimate is reported. The included controls should be justified and explained. For example, what is the year variable intended to capture? The specification seems fairly sparse compared to other meta-analysis regression analyses. Are there other controls that could be included? Some possibilities include: whether the given data set was used, whether the studies focused on the same or a similar population, whether a given researcher was a co-author.

Alternative tests

To address the issue, mentioned above, of equal weight given to multiple estimates from a paper, we suggest exploring whether the mean of group means estimator could be blended with estimator 3 and with estimator 4 which take into account information on sample size and sampling error variance. In the first case, estimator 2 could be modified so that each group mean is weighted by the average sample size for study j .

An alternative approach to the estimates in 1-5 is to compute the median across studies and see if this value lies within the confidence interval for the individual estimates. This approach is robust to outliers, but fails to make use of other information.

These estimates assume no correlation in the VSL estimates across studies. Following Hsiang et al. (2013), one could assume positive correlation across such studies and investigate if there is a consequential effect on the variance of the average estimate.

Can the parametric model (11) be estimated with weighted least squares? It appears to have the same structure as a standard random effects model. If not, the reason why the simulated likelihood approach, which requires additional structural assumptions, is needed should be explained.

The authors could conduct a publication bias test, following Card and Krueger (1995). The basic idea is to look at t-stats on the estimate of interest and see if it is positively related to the sample size (or degrees of freedom). If researchers are selectively reporting results, then there should be no or a negative relationship.

References

Hsiang, S.M., Burke, M., and E. Miguel. 2013. Quantifying the Influence of Climate on Human Conflict. *Science* 341 (2013): 1235367.

L. V. Hedges, I. Olkin, *Statistical Method for Meta-Analysis* (Academic Press, 1985).

D. Card, A. B. Krueger, Time-series minimum-wage studies: A meta-analysis. *Am. Econ. Rev.* 85, 238
(1995).

Dr. Kerry Smith

Draft Responses to Charge Questions EEAC Review of VSL White Paper

V. Kerry Smith

Preliminary draft (intended for discussion only)

02/29/2016

1a. Evidence of Validity of Stated Preference Studies

I reviewed the February 2016 draft white paper and the text in Appendix B describing how the specific estimates from stated preference studies were selected. I did not find a clear set of criteria for inclusion. Table 2.1 on page 13 of the Robinson–Hammitt [2015] paper lists a set of selection criteria for both the revealed and stated preference studies. The 10th item calls for evidence of validity of the estimates “. . . including sensitivity of willingness to pay to changes in risk magnitude.”

I did not find specific information, aside from responsiveness to risk, that would be used. Thus, this aspect of the white paper and the Robinson–Hammitt paper was not developed. Perhaps the foremost omission is the issue of incentive compatibility of the CV questions. Since Carsons and Groves [2007] we know a format for contingent valuation questions that is incentive compatible. Carson, Groves and List [2015] extended the initial arguments and demonstrated their importance. It is difficult to imagine why this discussion was omitted. I suspect it may be that few of the CV studies followed this format. If that is the case, then we need to have a discussion considering the potential bias when it is not followed.

In addition, the literature on stated preference (SP) studies would identify indirect measures of consistent pattern such as:

- a) evidence of understanding the choice asked in the SP question
- b) evidence of a statistical relationship between variables that economic theory suggests should influence the stated choices (the risk change is only one of the possible measures)
- c) evidence from pre-testing, focus groups, pilot surveys, etc. of evaluation of survey materials consistent with (a).
- d) examination of item nonresponse patterns that would indicate lack of understanding
- e) evidence of inconsistency in response to questions of perceived risk or stated adaptations to risk in respond to proposed activity that do not match the choices used for valuation estimates

These would be in addition to what the white paper refers to as a “weak scope” response to the risk changes that are asked about in the survey. Proportionality is not explained in the document. I was not sure if the proportionality referred to the WTP measure derived from the CV response or the VSL estimate. The former would be more consistent with other uses of contingent valuation. The latter would be more sensitive to the model used to define expected utility and the VSL as explained below. A citation would be useful.

Using the Eeckhoudt – Hammitt [2001] formulation for the expected utility model underlying the VSL one would have $A = \text{alive}$, $D = \text{Dead}$

$$U_A = U_A(m) \quad m = \text{wealth}$$

$$U_D = \alpha U_A(m) - \delta \quad 0 \leq \alpha \leq 1$$

$$1 - p = \text{probability of survival} \quad \delta \geq 0$$

In this framework, VSL is defined as

$$VSL = \frac{(1 - \alpha)U'_A(m) + \delta}{(1 - p + p\alpha)U'_A(m)}$$

$$\frac{\frac{\partial VSL}{\partial p}}{VSL} = \frac{(1 - \alpha)}{1 - (1 - \alpha)p}$$

The point is that the expectations about how VSL changes with p depend on the structure of the model underlying the derivation of the VSL.

This relationship is different from how the scope test was originally formulated by the NOAA Panel where a non-market service was assumed to enter individual preferences, but analysts did not know how amounts of the service should be measured. As a result clarification in what the weak scope test is and how it relates to the structure of the model used to define VSL is warranted.

It is important to acknowledge that information associated with the questions listed above usually is not reported in journal articles. As a result EPA should seek to obtain a copy of the questionnaire, any background testing of survey instrument, and information that would address these questions if it is not part of the published article and/or online material.

Estimates should not be accepted as valid simply because they are published.

1b. Construction (or selection of the Measure for the Risk Variable used in Hedonic Wage Models

The issue implicitly raised in the preference for risk measures based on occupation and industry is a maintained assumption that other measures for risk that are distinguished at the industry level are subject to serious errors in variables issues.

Black and Kniesner [2003] considered alternative risk measures and their conclusions were much more nuanced than the discussion here. Based on discussion of the potential for

endogeneity of the risk measure (see Cropper et al. [2011] pp 320-321 for a summary) together with the Black and Kniesner paper and early discussion of errors in variables and simultaneity in Garen [1988], this implicit assumption that the use of CFOI data resolves the matter seems incomplete. It needs further discussion and evaluation and may not be an appropriate basis for screening out papers that did not use risk distinguished by both occupation and industry.

1c. Estimates for Immediate Risk Reductions

There is extensive research on risk communication and long latency risks that suggests the typical respondent has difficulty interpreting the risks. (Cropper et al. [2011] cite two CV studies with examples of these issues.) Given the research indicating that lay persons misinterpret long latency risks, it would seem imposing a discounting assumption is not warranted without some empirical evidence supporting the practice.

Would these estimates be judged as outliers in the meta regression using conventional regression diagnostics?

2. Relevant Empirical Studies Included

The screening criteria based on CFOI and the use of industry and occupational risk measures need to be reconsidered in light of the comments under (1b). This reconsideration could allow a larger set of hedonic wage studies to be included.

A second consideration arises in the criteria for including studies (general criteria, point 3) requires that the study be based on a sample of the general U.S. population. This requirement is misleading for hedonic wage studies for several reasons:

- a) The hedonic wage model is only relevant to individuals who are working; most of the analyses do not include the decision to work; thus this is a subset of population.

- b) Available samples CPS or otherwise are subject to non-response effects based on who responds to the sample; the PSID began with one criteria and has evolved over time. There is not a specific discussion of these considerations in this criteria. Similarly, the Knowledge Networks panel is argued to be representative of the U.S. population, but these respondents may display special unobservable features that attract individuals who will agree to complete a specified number of internet based surveys per month on exchange for internet services.
- c) The wage hedonic model is based on a maintained assumption of a market equilibrium that gives rise to the function. The estimated wage/risk tradeoff is often treated as a “local” constant relevant to non-workers. It would be desirable to have a more detailed discussion of the relevance of tradeoff measures from samples designed to be representative of populations that are a subset of the U.S. population and then consider controls as part of the meta analyses of published estimates.
- d) The selection criterion calling for the use of samples of the U.S. population appears to contradict the proposal to use weights (discussed in charge question #3 and Appendix B).

3. Weighting Sub-Population Estimates

Answers to this question require that one distinguish how the weights are used in different sets of studies, so each will be considered separately.

- a) Cameron – De Shazo studies. These three studies appear to be based on their overall contingent valuation survey. To judge the weights more information on the authors’ sample design would be desirable. Assuming it was designed to be representative of non-institutionalized, English speaking adults, it does not appear there was an adjustment for English speaking in the construction of the weights for EPA’s summary. I am not sure if

the survey was designed for non-English speakers, so this adjustment could be important.

This comment applies to weights for all the Cameron – DeShazo studies.

- b) The Viscusi – Aldy (2007) study uses the age group specific estimates of VSL by the proportion of the age group in the entire population in 2013. There are several issues here. The estimates are for 1998 not 2013. Second, they exclude agricultural workers, members of the armed forces, those with less than ninth grade education, workers with less than minimum wage and non-fulltime employers. The weights used to represent the population that the sample is hypothesized to represent. The proposed weighting is not based on this criterion. This strategy mixes the assumptions one makes for benefits transfer with those associated with developing a population estimate. Each should be discussed and identified separately.

The empirical estimates are not relevant to the age groups in the whole population (except by assumption; and this should not be buried in weighting assumptions). There is also the issue of non-English speakers and how they were treated in the original 1998 CPS.

- c) Viscusi – Aldy (2008). These data are for 1993–2000; they exclude agricultural workers, members of armed forces, workers making less than minimum wage, less than fulltime workers and those with top coded incomes. Again the population in 2013 is the wrong group—weights for the period of the sample and the group represented in the sample (not the whole population 18-62).
- d) Viscusi – Hersch (2008). Again this analysis is for workers in 1995–96, 1998–99, and 2001–2002. Using the 2012 proportions of adults who smoke does not match the sample design. Since the decision to smoke has been used as an indicator of attitudes toward

risk, and these have changed over time, I can see in this case why one might want to use current fractions of smokers. However, the perception of the risks and degree of risk aversion of current smokers in relation to nonsmokers may have changed as well.

Indeed, Viscusi's book *Smoking: Making the Risky Decision* (Oxford 1992) documents that smokers tend to understate smoking related risks (at least from the surveys associated with the book which are now 25 years old).

All of these factors suggest I am not sure how these estimates should be treated. Regardless of when (the year) the distinction between smokers and nonsmokers is measured, the relevant base corresponds to groups that the sample is intending to represent, not the full population of the US!

4. Computing Standard Errors

If the weights are adjusted as discussed under #3 then the methods for estimating standard errors need to be adjusted to reflect the weights used.

The procedures for computing standard errors for the VSL must consider how the wage was estimated. It was difficult to determine how this issue was handled in Appendix B. If the wage is estimated using the estimated hedonic function as a conditional prediction, then the bias in using $\tilde{W} = \exp(\widehat{\ln W})$ and the var $(\widehat{\ln W})$ need to be considered.

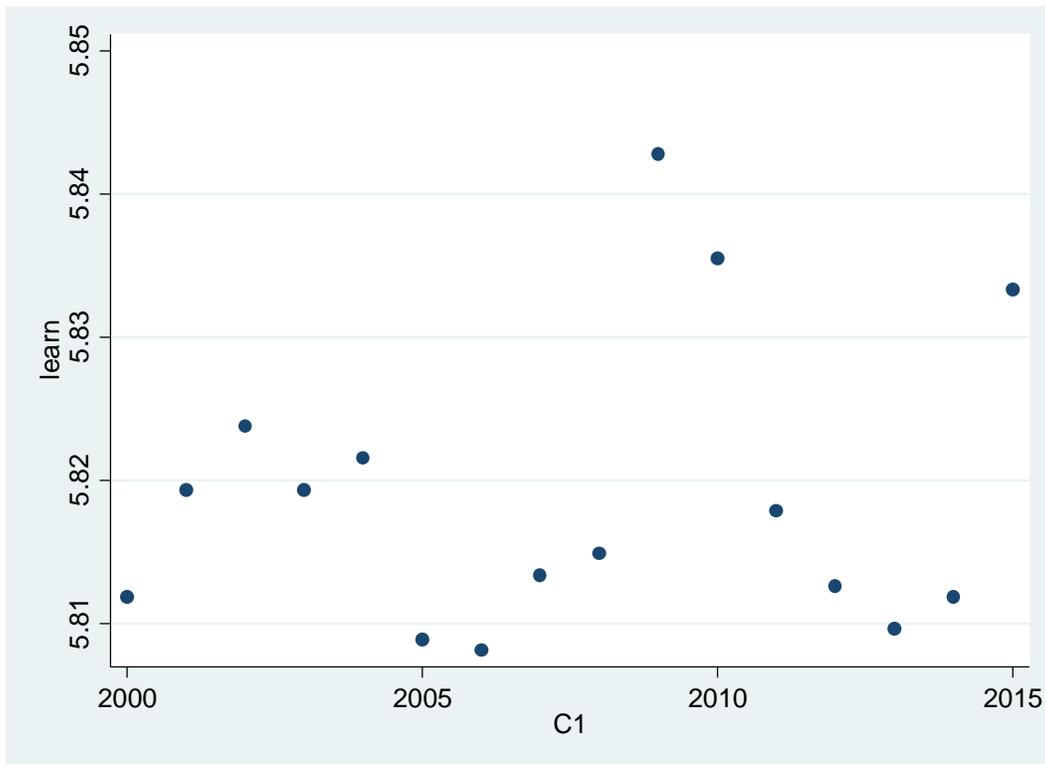
5. Application of Meta Analysis to VSL Estimates

The White paper's approach to the meta analysis is interesting and innovative. However, there are several issues that would be subject to question.

- a) It appears from the results reported in Table 6 and the Matlab code that the VSL estimates were adjusted for income growth using an assumed value of 0.7 for the income

elasticity and the index of income growth. The income growth measure appears to be what is labeled as the gdp factor. Since the risk tradeoff measure is estimated at the household level, the first question to raise is why use a gdp based measure?

The growth rate summarized in Robinson and Hammitt [2015] Figure 4.2 would suggest median weekly earnings from the BLS from 2000 to 2014 have fluctuated around 0% growth with negative values in 2005, 2010, 2011, 2012 and 2013. These would appear to offset the positive rate around 2009. The simple regression (log average earnings as a function of time – 2000-2015) and associated plot suggest no trend. Hence the growth assumptions seem unwarranted. The results from a simple analysis using Stata are given below:



Learn =log(weekly earnings)⁶

C1 =year

reg learn C1 if C1>=1999

Source	SS	df	MS			
Model	.000196037	1	.000196037	Number of obs =	17	
Residual	.001813906	15	.000120927	F(1, 15) =	1.62	
Total	.002009942	16	.000125621	Prob > F	= 0.2223	
				R-squared	= 0.0975	
				Adj R-squared	= 0.0374	
				Root MSE	= .011	

learn	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
C1	.0006932	.0005444	1.27	0.222	-.0004672	.0018536
_cons	4.426607	1.092648	4.05	0.001	2.097683	6.755532

Also while the Matlab code allows for sensitivity analysis associated with income elasticity, why impose this assumption on the estimates prior to a meta analysis? Why not consider testing the relationship between estimated VSL measures and alternative income measures? A simple regression analysis that first estimates conditional mean VSL for each study and then uses these estimates together with median household income or weekly earnings indicated no consistent relationship.

- b) There does not appear to have been consideration of the features of the studies as potential determinants of the VSL estimates. Separate estimates for hedonic wage and stated preference measures were considered; studies were dropped selectively as part of

⁶ One would also want to investigate changes in weeks worked and in annual earnings. I suspect over this period the annual earning would yield similar results but this should be verified.

considering influential studies, but there did not appear to be an effort to consider descriptive features of each of the studies as determinants in the meta analysis. Perhaps the reason was the effort to construct national average measures for VSL from each study. Nonetheless, it would be useful to consider why these additional variables were not evaluated in ways consistent with much of the literature.

This type of micro analysis could be treated as a “first step” in evaluating the role of specific estimates for the overall summary statistics. In this context this approach could serve as a complement to the methods used to estimate different types of means (as described on pages 22-25).⁷

6. Classifying Estimates into Groups

Footnote #11 on page 20 describes the groups. Some of the classifications do not appear to match what is given in Appendix B. For example, footnote 11 suggests that Aldy and Viscusi (2008) examined eight samples including the sample examined in Viscusi (2003) and others.

On page 60 of Appendix B, the descriptions of Aldy and Viscusi (2008) and Viscusi (2003) are not complete enough that a judge. Especially if one compares the criteria for excluding workers in some of the other categories. Perhaps the table summarizing the definitions of each sample as part of an appendix would help.

Studies relying on the same underlying data do not represent independent contributions to the composite mean estimate across this type of sample. It would seem one would want to account for the within group correlations (due to being based on the same samples). None of the

⁷ The results in Table 9 report a simplified version of this type of parametric model with controls for year of data collection, use of stated preference and whether the measure was a median or a mean. Based on the summary, I assume Y_{ij} is the VSL adjusted to be in 2013 dollars including the adjustment for income growth.

proposed means would allow for this correlation. Without the primary data from each study it is not clear it would be possible to account for the correlations.

This inability to account for the correlation should be noted.

7. Sampling and Non-sampling Errors in Optimal Weights

This question is conditioned by the model used to describe the data generating process, namely:

$$Y_{ij} = Y + \eta_i + \mu_{ij} + \epsilon_{ij}$$

Y is the true mean (assumed to be one value)

η_i – group level (i) non-sampling error

μ_{ij} – observation level non-sampling error

ϵ_{ij} – observation level sampling error

η and μ are assumed to have constant variances. Suppose we argued that η_i is better partitioned by method (SP versus HW) as well as group. This would imply a distinction that is evaluated through sub-sampling. However, we could also distinguish mode of questioning within SP, or source of risk data as an alternative to dropping those studies that do not use both industry and occupation in assigning risk. Were these alternatives considered? If so some discussion of the results would be desirable.

Bottom line – it would be useful to have some discussion of what the concept of “the group” means and its full implications for the various mean statistics considered. It appears to approximately correspond to the specific study based on the analyses that were undertaken here.

8. Other Approaches

A variety of analytical models have been used to bound the income elasticity of the VSL in relation to the coefficient of relative risk aversion. The technical memorandum describes some of these. It is also possible to derive a measure of the VSL from the labor supply elasticity. The prospect of using the CRR and/or labor supply elasticity as a cross check or bound for the VSL has not been considered.

The 2010 white paper noted:

“We recommended conducting additional scoping studies and further research to develop structural benefit transfer functions, possibly based on a life cycle assumption framework . . .” (P 60).

Intermediate efforts considering the possible use of literature reviews on the available estimates for CRR and labor supply elasticity that could be used to match the VSL estimates and provide a basis for cross checks to the meta summaries would seem consistent with this recommendation of the 2010 white paper. Based on what is reported in the white paper, Robinson and Hammitt [2015] and the technical memo on IEVSL these aspects of the 2010 white paper have apparently not been considered.

9. Preferred Estimates

The proposed summary measure (nonparametric, balanced mean of the study mean and parametric, balanced) seem like defensible estimates. However, there are important qualifications.

- a) These estimates have used weights to construct “general population” measures for the U.S. population. These details were questioned as part of my comments on the weights. As indicated in those comments, the weights do not seem warranted. It would appear more transparent to weight based on the population each sample is attempting to

represent and then explicitly identify the assumptions being made when the estimate is assumed constant for the whole US population. This later step is a benefits transfer.

Weighting component estimates based on the fractions of the US population assumes the estimates from studies are relevant to these sub-groups. For the hedonic wage models the estimates for risk/income tradeoffs were not intended to be relevant to the US population—they related to the workers covered by the specific sample underlying each model's estimates.

- b) Adjustment of VSL estimates by an income elasticity of VSL and index of income growth (especially one based on GDP per capita) does not seem appropriate. Conversion of VSL to current dollars would be appropriate. “Building in” the income elasticity and growth assumptions as maintained hypotheses before constructing the mean does not seem appropriate. Once again this is a part of the benefits transfer logic.

At present the documentation in the white paper is not clear. Table 6 of the text refers to the use of an income elasticity of 0.7. It is not clear on the source of the income growth index. The Matlab code seems to use gdp per capita. Footnote #19 in Appendix B refers to the estimates being developed using an income elasticity of 1.0. However it also notes that estimates will be replaced with VSL adjusted for price effects only before the SAB review.

Finally, footnote #17 refers to estimates of the income elasticity from 0.1 to 1.7 and alternative balanced means for the group mean as discussed in Appendix C. Thus the specific assumptions need to be clarified. I also feel that adjustment for income growth is not appropriate in constructing the means. It is a part of the assumptions that need to be

explicitly discussed in the benefits transfer component of the development of VSL measures for policy.

10. Influence Analysis

An influence analysis is important, especially given the implicit assumptions underlying the structure of the non-sampling error related to groups. It would be useful to consider the potential for regression diagnostic indexes (Belsley et al. [1980], Belsley [1991], and Cook and Weisberg [1982]) for the parametric modeling of VSL. These statistics allow analysts to consider if specific observations were influential to individual coefficients in the meta regression function. They allow an assessment of whether the magnitude and significance of individual coefficients was influenced by particular observations.⁸ Since these correspond to the specific studies and models within a study, they could help in understanding how the group definition noted earlier influences the specific mean statistics proposed to construct a population level measure for the mean VSL.

11. Regular Updates and Standardized Protocol

The VSL is based on a local estimate of marginal rate of substitution. Because it is derived (in the case of hedonic wage models) from a reduced form relationship that is assumed to approximate a market equilibrium, one should expect the factors that influence the ways individuals would tradeoff wealth for risk or market conditions, would be important to these tradeoffs. Examples would include: new regulations that affect risks or conditions of work,

⁸ These are “old” references but can provide useful indexes of how specific observations influence results. The discussion of “short data” in chapter 7 in Belsley may be especially relevant to parametric models developing meta summaries with limited variation in the risk and/or income measures that are used to estimate income elasticities or scope effects.

macro shocks to labor markets, major changes in national health policies, and new sources of environmental risks.

To move beyond this general guidance is in part a research task. Before discussing some specific suggestions, it is important to provide some context. The VSL is very likely the most important “benefit measure” used in EPA’s benefit cost analyses for air regulation and any other policy related to mortality risk. Indeed, the coal rule while motivated by climate policy had important ancillary benefits that depend on mortality risks. As a result, the level of staff effort and research resources devoted to regular updates should be commensurate with the importance of the estimates for policy evaluation. In my opinion it should be a high priority.

Some responses to the earlier questions affect the protocol for updates. I have divided my comments into two parts:

I. Research.

- a) We do not know the importance or impact of the Carson–Groves incentive compatibility criteria for SP estimates involving estimates of risk tradeoffs. The ability to use existing and past SP studies depends on whether the bias from failing to use incentive compatible questions is important for risk related questions. The ability to use fairly current research findings depends on an assessment of this issue.
- b) The existing protocol of screening hedonic wage studies relies on the assumption that using risk measures distinguished by industry and occupation resolves the errors in variables issues. It also seems to ignore sorting and endogeneity issues. Black and Kniesner initiated research comparing risk measures; this should be continued.
- c) There are a variety of economic index numbers used for policy. One of the most important is the Consumer Price Index (as well as other price indexes) used for cost-of-

living, social security, and other automatic adjustments. The CPI index is based on weights for categories of consumer goods. What are the protocols for updating the weights, adjusting for new goods, reflecting quality changes in existing goods, etc? Are there well-defined criteria for periodic updates? What are the standards? A comparison of the economic rationale for each with the context for policy uses of the VSL would be desirable.

II. Specific Suggestions

a) Evaluate using a panel and repeated cross sections of wage hedonic models to investigate how the income (or wealth) / risk tradeoff changes with:

- time
- extent of recession
- labor market conditions (e.g. unemployment)

by characterizing each cross section, pooling samples and testing for differences in the measures of tradeoffs for well-defined conditions that might be associated with a shift in the hedonic wage function.

b) The parametric model (in Table 9 of the white paper) reports how the VSL changed with year the estimated effect for the HW model was significant and positive. This might imply an increase in the tradeoff over time. However, it seems that the VSL estimates were adjusted using the income growth term and an income elasticity. As a result the positive effect could be due to the adjustment.

In some preliminary regression estimates I re-ran the analysis without adjustment.

- A. I estimated fixed effects for each study so the coefficient estimates (adjq1) would be conditional means of the VSL for each study (without income adjustment).

B. I used these coefficients for the fixed effects together with a variable for the year of the data (adjx5), a dummy variable for the HW studies (adjx4) and an interaction variable (int_adjx5 = adjx4 * adjx5).

In this content, the year variable and interaction did not indicate a significant change over the period of the sample (1996-2009). The results are given below. There were 18 studies but only 17 non-zero fixed effects could be estimated.

```
. reg adjq1 adjx1 adjx5 adjx4 int_adjx5 if adjq1~=0, noconstant;
```

Source	SS	df	MS			
Model	1398.15452	4	349.53863	Number of obs =	17	
Residual	160.513032	13	12.3471563	F(4, 13) =	28.31	
Total	1558.66755	17	91.6863265	Prob > F =	0.0000	
				R-squared =	0.8970	
				Adj R-squared =	0.8653	
				Root MSE =	3.5139	

adjq1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
adjx1	-925.4355	743.8739	-1.24	0.235	-2532.477	681.6062
adjx5	.4659381	.3714257	1.25	0.232	-.3364784	1.268355
adjx4	223.6595	1054.789	0.21	0.835	-2055.073	2502.392
int_adjx5	-.1098886	.5271738	-0.21	0.838	-1.248778	1.029001

```
. test _b[adjx5] +_b[int_adjx5]=0;
```

```
( 1) adjx5 + int_adjx5 = 0
```

```
F( 1, 13) = 0.91
Prob > F = 0.3586
```

This issue needs to be investigated more systematically.

C. It would be desirable to repeat the influence analysis discussed in the white paper and summarized in Table 10 but in this case omit studies with specific sample years.

12. Cancer Differential

The literature does not support a non-zero cancer differential. At this stage there is not sufficient research. All of the stated preference studies would fail the Carson–Groves criteria for incentive compatible responses. The literature does not offer sufficient resolution to allow one to judge the effects of the question formats used on the tradeoff estimates. Responses to earlier questions provide more detail on the issues involved.

13. Robinson–Hammitt (2015) Summary, EPA Technical Menu on the Income Elasticity of VSL and Income Elasticity of Nonfatal Health Effects Literature

I was unable to get access to the Viscusi [2015] paper. It appears to be in the *American Journal of Health Economics* (based on Robinson and Hammitt). I did review the NBER working paper version (WP# 20116, May 2014). Table 4, Panels A and B report results from meta regressions with income measures. The results do not correspond to the summary given in the technical memo's summary. I suspect there were revisions to Viscusi's paper before it was published but could not verify the conjecture. The statistically significant estimates for the income elasticity range from 0.409 to 2.272. Thus, it is difficult to reconcile these findings with the summary in Table 1 of the IEC technical memo.

More generally, the technical memo does not deal with the technical issues raised in Evans and Smith, ES, (JRU, 2010) which offer reasons why the income elasticity would be lower than implied by the coefficient of relative risk aversion. The memo summarizes the ES results but does not consider how they could be used as a screening criterion for selecting an income elasticity estimate.

Viscusi (Monthly Labor Review, October 2013) attributes the differences in the income elasticities from earlier studies that range from 0.5 to 0.6, versus the newer research using the CFOI to the improved risk information. However, as I cited, the 2014 Viscusi working paper

with the same title as the 2015 AJHE version of the paper, cites estimates ranging from 0.409 to 2.272. The sample for the 2014 working paper used the CFOI risk measures. A clear indication of the sources of the differences between the published and working paper version of the analysis seems warranted.

More generally, as noted earlier, the measure of income used may well be important. GDP per capita measures are not the relevant basis for assessment of the IEVSL. VSL is an individual tradeoff and an individual or household income measure seems more relevant. It should also be a measure that is likely to be relevant to the workers whose tradeoffs are represented by the VSL measures from the wage hedonic models.

Overall then, I did not feel the technical memo and Robinson–Hammitt provided adequate consideration of the link between theoretical expectations for the income elasticity and available evidence—including reconciling CRR and labor supply linkages.

I also found that it was difficult to reconcile the consensus measures with the estimates I could find in the literature these documents cited.

14. Stated Preference Estimates of the Income Elasticity

The white paper does a nice job displaying the sensitivity of results to the use of VSL estimates from HW and SP studies. I do not feel there is a basis for setting a threshold for estimates of the income elasticity, aside from suggesting that they should not be negative.

Sensitivity estimates paralleling the discussion of HW and SP in the white paper would be desirable. In addition, extending the discussion to consider what estimates for the CRR and labor supply elasticity would imply for plausible ranges of the income elasticities would serve to complement the analysis based on meta analysis. It is also broadly consistent with the recommendations of the 2010 white paper.

15. Robinson and Hammitt Selection Criteria

As noted in response to earlier comments, I did not feel relying on the use of risk measure distinguishing industry and occupation resolved the errors in variables, sorting, and potential endogeneity issues raised in the literature. I would have preferred to see a more detailed discussion, along the lines of Black and Kniesner of the advantages and disadvantages of risk measures that might allow admission of a larger set of studies for meta analyses.

As to the SP criteria there was no consideration of the Carson–Groves analysis of incentive compatibility of contingent valuation questions. A careful assessment of what is known about the bias of alternative non-incentive compatible methods would be essential to amending the selection criteria for SP studies.

The online material does not support including Viscusi, Huber and Bell.

16. Appropriate Criteria for Income Elasticity

My comments to earlier charge questions implied I felt the criteria needed to be re-evaluated.

There is a more general issue to be raised. Using measures of income considered relevant to HW VSL estimates, there has not been consistent household income growth for the groups that are best represented by the VSL estimates since 2000 in the United States. In light of this evidence and: (a) the limited nature of the available estimates for the IEVSL; plus (b) the need to do further research to reconcile the estimates of the IEVSL with established economic theory linking income elasticities, CRR and labor supply, it might be prudent to suspend adjustment to the VSL for hypothesized income growth. This strategy implies VSL would be adjusted for changes in the price level only until research was completed on the issues associated with screening criteria for IEVSL estimates and sustained household income growth was observed.

17. Income Elasticity for value of reducing risks of Non-fatal Health Effects

There is no basis for assuming that the income elasticity for non-fatal health effects corresponds to the IEVSL. Moreover, the magnitude of the valuation estimates and the income elasticities seem likely to be directly impacted by the changes in national health insurance policies. Until there is a clear theoretical assessment of the linkages between these concepts for the income elasticity as well as a better understanding of the role of health insurance policies for these tradeoffs, this strategy cannot be justified as an appropriate and scientifically sound approach for specifying the income elasticity of the value for reducing the risk of non-fatal effects.

References

Belsley, David A. 1991. Conditioning Diagnostics: Collinearity and Weak Data in Regression.

(New York: John Wiley).

Belsley, David A., Edwin Kun and Roy E. Welsch, 1980. Regression Diagnostics: Identifying

Influential Data and Sources of Collinearity. (New York: John Wiley).

Black, Dan and Thomas J. Kniesner, 2003. "On the Measurement of Job Risk in Hedonic Wage

Models." *Center for Policy Research*, Paper 181.

Carson, Richard T. and Theodore Groves, 2007. "Incentive and Informational Properties of

Preference Questions." *Environmental and Resource Economics*, 37 (1): 181-210.

Carson, Richard T., Theodore Groves and John A. List, 2014. "Consequentiality: A Theoretical

and Experimental Exploration of a Single Binary Choice." *Journal of the Association of*

Environmental and Resource Economists, 1 (1/2): 171-207.

Cook, R. D. and S. Weisberg, 1982. Residents and Influence in Regression. (New York: Chapman and Hall).

Cropper, Maureen, James K. Hammitt and Lisa A. Robinson, 2011. “Valuing Mortality Risk Reductions: Progress and Challenges” Annual Review of Resource Economics, edited by G. Rausser, V. Smith and D. Zilberman, (3): 313-336, (Palo Alto: Annual Reviews, Inc.).

Eeckhoudt, L. R., and Hammitt, J. K. 2001. “Background risk and the value of a statistical life” *Journal of Risk and Uncertainty*, 23(3), 261–279.

Garen, John, 1988. “Compensating Wage Differentials and the Endogeneity of Job Riskiness.” *The Review of Economics and Statistics*, 70 (1): 9-16.

Dr. George Van Houtven

PRELIMINARY RESPONSES TO CHARGE QUESTIONS

George Van Houtven

3/3/16

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.

I believe that the use of scope test results (including a weak scope test) as the main validity criterion is an appropriate one. However, the application of this criterion for the Viscusi et al (2014) study is unclear. In short, I think it passes the validity test but not for the reasons suggested in the paper. The White Paper (p.56) states that “the authors do not explicitly address tests. However, regression results indicate respondents are willing to pay more for treatment that reduces risk to zero.” First, I believe that by finding a statistically significant positive effect of Risk in the probit analysis, they have explicitly addressed scope. Second, I do not believe that a premium on reducing risk to zero is a strong indicator of validity. The Hammitt and Graham (1999) paper raises additional validity issues, in that the scope test is not significant for the subsample who correctly understood that 1/10,000 is greater than 5 in 100,000; however, this by itself should not disqualify the selected VSL values for the meta-analysis.

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.

Based on the information provided, the risk construction criteria used to include/exclude VSL from the hedonic wage studies appears to be appropriate.

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.

The inclusion of the Viscusi et al estimates is problematic for the stated reason. There is ample evidence from the literature that latency matters and that future risk reductions should be discounted, but little consensus thus far on what that rate should be. The selection of a 3% rate is not unreasonable given the current evidence, but it is a bit arbitrary. Given that multiple VSL values were included from the Cameron and DeShazo using discount rates ranging from 3% to 7%, it might make sense to then also include the Viscusi et al VSL estimate that uses a 7% discount rate. The Alberini study is less problematic because it asks for an annual payment corresponding to a concurrent risk reduction over 10 years.

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.

I am not aware of any other studies that meet the selection criteria and therefore should be included in the White paper.

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.

In general, the approach of using population shares to weight subpopulation-specific VSLs is reasonable; however, some of the applications could use additional explanation and justification. First, for Cameron et al (2013), the averaging by household size (Kids) in Table B-6 appears to assign equal weight to the 3 categories. Why was this assumption used? Second, for Cameron and DeShazo (2013) the approach for developing and assigning weights to the different subsample VSLs is unclear.

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.

For the stated preference studies, standard errors (or 95% confidence intervals) for the selected VSL estimates were apparently reported in all cases; however, additional work was often needed to get them into the proper form for the meta-analysis. In one case (Cameron et al., 2013) they were reported in a figure and therefore required a visual approximation of the relevant points. This approach is not ideal but is better than the alternative of excluding these estimates. In other cases, the approaches appear to be appropriate but would benefit from additional explanation. For example, in several cases the standard errors of the *mean* VSL needed to be estimated using a log transformation based on the reported standard errors of the *median* VSL. For completeness, the transformation approach for the standard errors should be explained in more detail. Currently, it is only described as “analogous” to the transformation of the medians (p.47). Similarly, standard errors were estimated for all of the *weighted* average VSLs from Cameron et al studies, using the standard errors from the individual sub-population-specific standard errors. The approach and assumptions (e.g., zero covariance) for calculating these weighted average standard errors should be explained.

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.

The methods used in the White Paper to analyze VSL estimates from the literature are for the most part scientifically sound and consistent with standard and accepted practices for conducting meta-analyses. However, to reinforce this conclusion it would be helpful for the paper to be more explicit about what these accepted practices are and how they are applied in the paper. This could be accomplished in a few ways. First, several papers have proposed general steps, guidelines, and/or recommendations for

conducting meta-analysis, with the most relevant being Nelson and Kennedy (2009). It would strengthen the paper if it were to organize the discussion around (or least reference) these types of best-practice guidelines. The White Paper does this to a limited extent with the PRISMA framework, but this really only applies to the study selection step. Second, the non-parametric statistical methods used in the analysis include approaches (“sampling error” and “total error” variance weighted mean) that are typically referred in the meta-analysis literature as “fixed effect size (FES)” and “random effect size (RES)” methods. Using, or least referring to, these labels would strengthen the presentation in the paper by tying it to the broader literature. The RES method is mentioned in the paper, but only in reference to the parametric/meta-regression approach.

When evaluated with respect to the 10 best-practice guidelines proposed by Nelson and Kennedy for meta-analysis in environmental and resource economics, the methods in the White Paper hold up well. However, the following issues should be considered and addressed.

- Standard tests of homogeneity across groups (Q-tests) are generally recommended, but they are not discussed or reported in the White Paper (with respect to the non-parametric models 4 and 5.)
- The fundamental error specification used in the analysis, which is summarize in equation (2), is essentially an RES structure, except that it includes one additional component – the observation-level non-sampling error. This expansion of the RES model seems reasonable, but it is not clear (to me) whether this is a novel meta-analytic approach or whether it has been applied in other studies. If the latter, provide study references. Either way the discussion should make this clear.
- The reason for including the sample size weighted mean method (number 3) is not clear. This approach is inherently inferior to variance weighted means and is typically used when variance estimates are not available, which is not the case for the VSLs included in this White Paper. Rather than including this method, a more interesting comparison might be to include a more standard RES structure that does not include a separate component for observation-level non-sampling error.
- The non-parametric analysis that is referred to as “mm”, which includes mean and median primary estimates in the same analysis, is not well justified and appears to be counter to best practices by including measures that not directly comparable. Moreover, it is not clear whether the same median and transformed median (i.e., mean) estimates are simultaneously included, which would certainly not be appropriate. A separate analysis that is median only would be more justifiable, but since the objective is to estimate mean VSL it is not clear what is gained by including medians.
- It would be interesting to know if there is a difference between the non-parametric model 5 (total error variance) and a parametric meta-regression model that only includes a constant term, since the underlying model structures are the same. Explaining this similarity or

difference would help to understand the connection between the parametric and non-parametric models.

- The reason for selecting the mean of group means as the preferred non-parametric VSL estimate deserves additional discussion. The justification is that it has the smallest standard error; however, this approach excludes information about the precision of the VSL estimates which seems very counter-intuitive.
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.

The approach for defining groups is appropriate; however, it could be more clearly explained. On page 20 they are defined as “independent data samples” with additional explanation in footnote 11. To make this group assignment more explicit, a column could for example be added to Table 6. For the purpose of sensitivity analysis, alternative groupings could be explored. For example, subsamples from individual studies could be grouped together into study-level groups.

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.

As discussed above, the approach for addressing different, hierarchical error components is appropriate and consistent with the FES and RES methods that are commonly used in meta-analysis. In one respect, it expands on these methods and offers an additional dimension by also including an observation-level non-sampling error component. If possible, it would be helpful to include and conduct a heterogeneity test that focuses on this particular error component.

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.

Yes, these approaches are generally appropriate, sound and defensible, except (as mentioned above) I see little justification for the combined mean-median (mm) non-parametric analysis, and I do not see the value in including the sample size weighted approach. For both approaches, I would be concerned if

the analysis includes both the reported median and calculated means from the same sample (i.e. Hammitt studies), which would be redundant (I could not see in the discussion if this is the case or not).

For the parametric analysis, although I understand that there are limited degrees of freedom, it would still be informative to see if and how differences in the size of the risk change matter, even if it is just a dummy variable that distinguished between those in the 10^{-4} and 10^{-5} range from those in the 10^{-6} range.

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.

Given the inherent differences between the hedonic and SP estimates and the fact that they each offer advantages and disadvantages for estimating VSL, the use of a balanced estimate that is an average of the two separate meta-analyses is appropriate. One adaptation of this would be to do a variance weighted average of the two, using the estimated standard errors for HW and SP models.

For the reasons mentioned above, I am less convinced about the use of the mean of study means from the parametric analysis because it does not include available information about the relative precision of the different estimates. Ignoring this information seems counter-intuitive, unless there are concerns about the accuracy of these precision estimates.

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 an appropriate approach and provides interesting insights. Most importantly it points to the sizeable upward effect of the Corso et al study on the summary VSL estimates. One possibility is that the calculated mean estimates, which are derived by EPA from the reported median values and their MSE, are inaccurate. For example the log transformation described on p46 may not be accurate if the errors of the original regression are not log-normally distributed (Duan, 1983). One additional sensitivity analysis that I would recommend is to exclude all of the “calculated means.”

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.

While it makes good sense to update the VSL meta-analysis on regular 5 year intervals, I believe that the decision about how far back in time to go to exclude studies should be made on a case-by-case basis, based on professional judgment about the relevance of the existing studies, rather than on a pre-determined basis.

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?

Even if the selection criteria were to be relaxed for analyzing the existence and magnitude of a cancer premium, I doubt that the number of defensible estimates in the literature is large enough to support a formal meta-analysis. Moreover the evidence is mixed at best regarding the presence of a cancer premium, and it is often difficult to disentangle them from morbidity and latency effects. At this stage, a systematic literature review and summary of the relevant literature would most likely provide more useful information for policy analysis than trying to conduct a meta-analysis.

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.

I am not aware of any other relevant studies.

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?

I do not agree with the logic of excluding mean zero estimates, unless there are other reasons to doubt the quality of the particular analyses generating those estimates.

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).

Yes, the Viscusi et al. estimates should be included.

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?

I would argue in favor of including the SP estimates. The fact that they are lower than expected should not disqualify them from consideration.

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.

Given the limited data available for non-fatal effects, this is a reasonable approach and it is consistent with the limited evidence that does exist for non-fatal effects (including the Van Houtven et al. [2006] meta-analysis referenced in the report.

References

Alberini A, Cropper M, Krupnick A, Simon NB. 2004. Does the value of a statistical life vary with age and health status? Evidence from the US and Canada. *Journal of Environmental Economics and Management* 48(1): 769-792.

- Chestnut LG, Rowe RD, Breffle WS. 2012. Economic valuation of mortality-risk reduction: stated preference estimates from the United States and Canada. *Contemporary Economic Policy* 30(3):399-416.
- Hammitt JK, Haninger K. 2010. Valuing fatal risks to children and adults: effects of disease, latency, and risk aversion. *Journal of Risk and Uncertainty* 40:57-83.
- Robinson, Lisa A. and James K. Hammitt. 2015. The effect of income on the value of mortality and morbidity risk reductions. Review draft prepared for U.S. EPA. Contract EP-D-14-032 with Industrial Economics, Inc.
- USEPA. 2011. Review of Valuing Mortality Risk Reductions for Environmental Policy: A White Paper (December 10, 2010). Office of the Administrator, Science Advisory Board. EPA-SAB-11-011. July 29. Available at:
[http://yosemite.epa.gov/sab/sabproduct.nsf/298E1F50F844BC23852578DC0059A616/\\$File/EPA-SAB-11-011-unsigned.pdf](http://yosemite.epa.gov/sab/sabproduct.nsf/298E1F50F844BC23852578DC0059A616/$File/EPA-SAB-11-011-unsigned.pdf)
- Viscusi WK, Huber J, Bell J. 2014. Assessing whether there is a cancer premium for the value of a statistical life. *Health Economics* 23:384-396. [On-line appendix available at:
<http://onlinelibrary.wiley.com/doi/10.1002/hec.2919/supinfo>]

Dr. JunJie Wu

Preliminary Written Comments/Response to

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”

JunJie Wu, Oregon State University

Charge Question #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.

Preliminary Comments on Charge Question #4:

- a. This is an extremely important question because the magnitude of standard error estimates is only indicator used to select the “preferred” models and non-parametric estimates of VSL.
- b. This question is classified as a question related to “Meta-analysis dataset,” but it is really a question about the “Analysis” itself because it asks for “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.”
- c. Given the important role that standard errors play in selecting preferred models and in determining the reliability of VSL estimates, the paper does not devote enough space to discuss the methods used to estimate standard errors. In fact, the paper includes only two paragraphs to discuss methods to estimate standard errors for non-parametric VSL estimates (section 4.1.1) and does not talk about the methods used to estimate standard errors for the parametric VSL estimates at all.
- d. Because it’s very brief, it is not quite clear about how the bootstrap estimation is done. For example, the paper states that “To maintain the within-group correlation structure among the observations, we randomly drew I sets of groups with replacement from the primary sample of grouped observations. We did not re-sample observations below the top (group) level (Davison and Hinkley 1997 p 100-101, Ren *et al.* 2010).” (p. 25). It’s unclear how each I set of groups were drawn and why re-sample observations below the top level were not re-sampled. In fact, it is unclear what is a “group/data sample”. In

footnote 11 on page 20, the paper states that “Hammitt and Graham (1999) and Corso, Hammitt, and Graham (2001) each examined 4 samples.” However, when looking at the last column of Table 6 on page 17, it seems that Hammitt and Graham (1999) examined only one sample and Corso, Hammitt, and Graham (2001) examined three samples.

- e. Related to the previous comment, there some notional problems. In Figure 1, N is used to denote the number of papers. But on page 20, N is used to denote “primary VSL estimates. It would help the reader if the authors could show in Table 6 a) number of samples/groups examined (I); b) the number of “primary VSL estimates” (N), and the number of observations from each group (m_i).
- f. The paper uses a bootstrap approach to estimate standard errors for non-parametric VSL estimates. However, there is an alternative way to calculate standard errors for each non-parametric VSL estimator. For example, when the “simply mean” estimator is used, the standard error of the estimate, by definition, equals

$$\begin{aligned} E[(\hat{y} - E\hat{y})^2]^{1/2} &= E\left[\sum_{i=1}^I \sum_{j=1}^{m_i} (\eta_i + \mu_{ij} + \epsilon_{ij})^2\right]^{1/2} \\ &= \sum_{i=1}^I \sum_{j=1}^{m_i} (\sigma_\eta^2 + \sigma_\mu^2 + se_{ij}^2)^{1/2}. \end{aligned}$$

Thus, once σ_η^2 , σ_μ^2 , and se_{ij}^2 are estimated, one can use the above formula to estimate the standard error of the VSL estimate directly, as an alternative to using the bootstrap approach. At least the paper should compare the results from the two alternatives because the direct approach is more consistent with the theory underlying the nonparametric approach.

- g. Currently the white paper does not provide any discussion about the approach used to estimate the standard error of VSL estimates for HW approach. I assume it uses the standard error of β_0 from the Hedonic wage equation regression as the standard error of the VSL estimate. Is this the right approach? Alternatively, because the Hedonic wage regression provides estimates for σ_η^2 , σ_μ^2 , and se_{ij}^2 , one can the standard error of the VSL estimate by

$$\begin{aligned} E[(\hat{y} - E\hat{y})^2]^{1/2} &= E\left[\sum_{i=1}^I \sum_{j=1}^{m_i} (\eta_i + \mu_{ij} + \epsilon_{ij})^2\right]^{1/2} \\ &= \sum_{i=1}^I \sum_{j=1}^{m_i} (\sigma_\eta^2 + \sigma_\mu^2 + se_{ij}^2)^{1/2}. \end{aligned}$$

Conceptually, this seems like a reasonable approach because the standard error of VSL estimates should be related to σ_η^2 and σ_μ^2 . In contrast, the standard error of β_0 reflects the variance of VSL estimates across the studies and has no direct connection to σ_η^2 and σ_μ^2 .

Charge Question #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).

Preliminary Comments on Charge Question #15:

Robinson and Hammitt (2015) is very well written, and provides a clear discussion of the selection criteria they used to select the studies. For the most part, Robinson and Hammitt (2015) use criteria that are consistent with those suggested in the 2011 SAB-EEAC review. However, Robinson and Hammitt (2015) do use a few criteria that are different from the ones suggested in the 2011 SAB-EEAC review. For example, Criterion 2, "be publicly available," is broader than the criteria suggested by the SAB-EEAC, who suggest limiting the search to the peer-reviewed literature. Robinson and Hammitt (2015) note that there are advantages and disadvantages of both approaches, and it's a judgment call which approach is better. Clearly, validation weighed heavily when the SAB-EEAC suggested limiting the search to the peer-reviewed literature. When an individual's judgment differs from a SAB-EEAC Committee judgment, and there is no clear proof that the former is better, it is my belief that it is more appropriate to subject individual judgments to the committee judgments, rather than the other way around.

Including Viscusi, Huber, and Bell (2014) in the Technical Memorandum would make it more consistent, at least in selecting studies, with the EPA estimates of VSL.

Charge Question #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?

Preliminary Comments on Charge Question #16:

There are pros and cons of two alternatives. However, there is a third alternative. Instead of putting equal weights on mean results from the HW and SP literatures, it might be more appropriate to explore alternative ways to select weights as EPA did when using non-parametric approaches to estimate VSL.

Some minor comments on the White Paper

p. 5. Start a new paragraph from “The specific choice of measurement units is arbitrary;...” The discussion of measurement units has no connection with the preceding discussion.

p. 7. Table 1. Insert “on” after “based”.

p. 20. Again, it is unclear what is a “group/data sample”. In footnote 11, the paper states that “Hammitt and Graham (1999) and Corso, Hammitt, and Graham (2001) each examined 4 samples.” However, when looking at the last column of Table 6 on page 17, it seems that Hammitt and Graham (1999) examined only one sample and Corso, Hammitt, and Graham (2001) examined three samples. It would be helpful if the authors could show in Table 6 a) number of samples/groups examined (I); b) the number of “primary VSL estimates” (N), and the number of observations from each group (m_i).

p. 24. In Equation (7), $var(y_{ij})$ should be $E[var(y_{ij})]$.

p. 28. Table 7’s footnote c. “Numbers is” should be “Numbers in”.