Meta-analysis of Mortality Risk Valuation Estimates
Draft Responses to Charge Questions

1. *In light of the workgroup's findings, what approach or approaches are the most scientifically appropriate to derive summary estimates of mortality risk valuation for use in environmental policy analysis? Should meta-regression techniques be applied to selected estimates or are other methods (e.g., fitting distributions) more appropriate? Please specify which methods, aside from or in addition to meta-regression techniques the Agency should explore.*

We believe that, once EPA has assembled a set of studies that are applicable to the population affected by a regulation and that meet appropriate criteria for a well-executed study, it is appropriate to combine these using meta-analysis. For example, the VSL estimates could be combined using a random effects estimator, in which individual VSL estimates are weighted in inverse proportion to their variance. We do not, however, believe that meta-regression is an appropriate way to combine VSL estimates, or to perform benefits transfer, for reasons described below.

A meta-regression, in which the characteristics of study participants (e.g., percent female, percent over 65) and study design (e.g., does an hedonic wage study control for a worker's industry) are used as covariates is useful in understanding what factors affect empirical estimates of the VSL. It can be thought of as an empirical literature review which may highlight factors affecting the VSL that would not otherwise be detected. By highlighting correlates of the VSL, meta-regression may suggest features of study design for which criteria should be established. For example, it might be determined that an acceptable hedonic wage study must control for the worker's industry at the 2-digit level because this has a significant effect on the VSL estimate obtained in an hedonic wage study

It is, however, another matter to treat a meta-regression as a reduced-form model that can be used for obtaining the VSL for a given sub-population or the VSL conditional on an appropriate study design. To illustrate, a meta-regression may control for the functional form of the dependent variable in an hedonic wage regression by setting a dummy variable equal to 1 if the dependent variable is the log of wage rather than the wage. If researchers believe that the appropriate form of the equation is to use the log of the wage, this should be one criterion for an acceptable study and only studies satisfying the criterion should be combined in the meta-analysis. Setting the dummy equal to 1 in the meta-regression is not equivalent to altering the functional form of the underlying studies.

Similar problems exist when population characteristics in the meta-regression are used for benefits transfer. Suppose that one of the covariates on the RHS of a meta-regression is the proportion of the study sample over 65. It is one thing for estimates of the VSL across studies to show that this coefficient is negative and statistically significant,

suggesting that the VSL is lower for persons over 65, and another to set this variable equal to 1 to compute the VSL for persons over the age of 65. The latter treats the meta-regression as a reduced-form model that can be used for benefits transfer. If the meta-regression suggests that age is important, EPA should use the results of studies that control for age, and other factors that are correlated with age, using individual data. These factors (e.g., wealth and income) are controlled for imperfectly in a meta-regression in which the population characteristics are summarized by an aggregate number for each study. The results of various studies may be combined in a structural model, but this is not what is happening in a meta-regression.

When meta-analysis (e.g., a random effects estimator) is used to combine studies, it is imperative that (a) the studies pertain to the population affected by the regulation; and (b) that the studies satisfy appropriate criteria regarding their design. We urge EPA to establish such criteria. In combining studies we do not recommend the use of quality weights, other than the 0-1 weights that are implicit in deciding which studies are of sufficiently high quality to be included in the meta-analysis. Although, in principle, there is no reason why weights should not vary between 0 and 1, in practice determining these weights is likely to be difficult. In the interests of transparency, we urge that a set of criteria for acceptable studies be established and then applied to the literature.

Formulating a list of criteria for an acceptable study and applying them to the literature will necessarily involve expert judgment. The committee believes, however, that EPA should not directly elicit appropriate VSL values from experts, asking them, for example, to specify a range of acceptable VSL values and/or a mean value based on their knowledge of the literature. This requires that the expert combine mentally the results of dozens of studies, and one loses transparency in the process.

2. *Using the approach identified above, what measures/estimates should be combined? VSL estimates? The coefficient on fatal risk? Other? How should the Agency select the measures to be combined? Should a single, preferred estimate be selected from each study or should all estimates be included?*

We believe that a meta-analysis (e.g., a random effects estimator) should be used to pool VSL estimates from acceptable studies that pertain to the population affected by a regulation. The committee recommends that only one estimate should be selected from a study that reports several models all estimated from the same dataset. Which estimate should be selected when a study reports several estimates of the VSL depends on the set of criteria that the Agency establishes to determine what is an acceptable study. For example, if different models use different sets of covariates, the Agency should select the model with the preferred set of covariates.

3. *Should original studies be required to use a common empirical specification (functional form and choice of covariates) in order to be included in a meta-analysis? What data are required of the original studies to be included?*

For compensating wage studies, the committee recommends that original studies report the results of a common specification of the compensating wage regression, in addition to the author(s)' own specifications. The compensating wage study report or article should also report the estimate of the VSL calculated by the authors and its standard error—the latter being essential for creating the weights to be deployed in the meta-analysis—and ample details on how exactly both were calculated.

We recommend that the compensating wage studies ultimately to be included in a meta-analysis:

- provide information on the source of data on risk, worker pay and worker characteristics;
- include codes for creating the sample used for the compensating wage regressions and for transforming variables;
- report average risk and average pay for the sample;
- explain whether the author(s) did or did not include non-fatal risks in the compensating wage regression, in addition to the fatal risks variable
- explain whether the sample contains only union workers, or if union membership was controlled for in the regression
- explain whether high-risk workers (e.g., police officers, firefighters, etc.) are included or excluded from the sample
- explain clearly whether the researcher(s) included a quadratic term in risk, and interactions between risk and other variables, in the regression.

For selecting stated preference studies, we recommend that the answers to the following questions be included:

- Was the study a (i) contingent valuation survey, (ii) conjoint choice experiments, or (iii) another type of hypothetical valuation exercise?

- What was the mode of administration of the survey?

- What was the sampling frame? Was a specific population targeted, or was the sample supposed to mirror the general population?

- Provide information about the age of the respondents, split by gender, income, education, health status (if available)

- What was the type of risk reduction respondents were to value in the SP study? (Was it a reduction in the risk of dying for all causes? Cardiovascular/respiratory causes? Road-traffic accidents?)

- Was the risk reduction immediate and incurred over the next year, or was it delayed into the future?

- Were respondents asked to consider a private risk reduction or one delivered by a public program? if public program, what was the payment vehicle (taxes, increases in the prices of products or the cost of living)?

- Was the payment one-time or an annual payment to be repeated each year for a number of years? How many years?

- Did each respondent have to value more than one risk reduction in the survey?

- Was the size of the (annual) risk reduction varied across respondents in the study? If so, (i) report the min, max and average risk reductions used in the study, and (ii) did WTP pass the scope test?

4. *Given the various approaches used in the literature, what is the most scientifically appropriate measure to derive when combining estimates from multiple studies? A single central point estimate, a single distribution, or a range of estimates in economic analyses? How can such a measure best reflect the uncertainty and variability in mortality risk valuation estimates?*

Meta-analysis should be used to provide a description of the probability distribution of the estimates. The resulting probability distribution can be used for uncertainty analysis, and the expected value and other relevant point estimates (e.g., median, 5th and 95th percentiles) can be drawn from it.

5. *How should stated preference studies and revealed preference studies be considered together in a scientifically appropriate method to derive summary estimates of mortality risk valuation?*

The committee believes that separate meta-analyses should be performed of stated and revealed preference studies. We advise the Agency to report separate valuation estimates based on stated preference and revealed preference estimates of the VSL. Our reason for this recommendation is that (a) the studies measure different concepts and (b) the economics profession may view them differently. Revealed preference (RP) studies, such as compensating wage studies, measure the rate of substitution between risk and wealth. Stated preference (SP) studies measure ex ante willingness to pay for a risk

change, and the resulting responses must be converted to a VSL. Regarding the second point, SP studies have been criticized because they rest on answers to hypothetical questions rather than on observed behavior. Furthermore, they often demonstrate respondents' difficulty in providing consistent answers to such questions. RP studies on the other hand rest on the maintained hypothesis that individuals correctly perceive risks as the researcher measures them—for example, that parents correctly estimate the reduction in risk of death or injury from a child wearing a helmet when riding a bicycle. The notion that people have good quantitative estimates of small risks makes some persons uncomfortable. It is also the case that compensating wage studies must infer the VSL by controlling for other factors that affect the wage, such as worker ability, which are difficult to measure and may be correlated with risk of fatal injury on the job. For all of these reasons, we believe that VSL estimates from RP and SP studies should not be combined.

6. *How should the Agency use studies based on specific sub-samples (e.g., elderly) in developing summary estimates of mortality valuation estimates for environmental policy analysis?*

As the committee noted in answering Life Expectancy Charge Questions 1 and 2, EPA should aim to distinguish the VSL according to age and, possibly, health status, the empirical correlates of life expectancy. This implies that separate meta-analyses would be performed for studies of different populations, for example, persons 30-65 and persons 65 and over.

7. *Most studies that combine estimates adjust the data from the original studies to some extent. For example, some studies adjust for after-tax wages, whereas others do not. Is there a set of such modifications that the SAB-EEAC believes to be critical when deriving summary estimates from the literature? Are there some data modifications that are generally incompatible with a sound approach to synthesizing existing estimates? What are the implications for interpreting results?*

In synthesizing estimates from multiple studies, it is important to adjust as well as possible for differences among the studies that have predictable effects on their results. Such adjustments can be made to studies before including their estimates in a meta-analysis.

One adjustment that can and should be made is to adjust for monetary inflation between studies by converting all nominal monetary values into real values. Because there is uncertainty about the best estimate of inflation over a period (reflected, for example, in different measures such as the various consumer price indices and the GDP deflator), the best adjustment is not clear. However, if one relies only on relatively recent studies (that

are likely to be most relevant to the evaluation of current policy), differences between the alternative indices are likely to be modest and contribute little to uncertainty about the appropriate valuation compared with other factors. Similarly, if valuation estimates using other currencies are included, it is necessary to adjust for the exchange rate (again, uncertainty about the best exchange rate is not likely to be a major concern).

Other adjustments that are in principle desirable are more difficult to make, and so the committee recognizes our current empirical abilities may not permit these adjustments. One is to adjust for differences in real income and wealth between study populations. Since the value of reducing mortality risk increases with income and wealth, differences in these factors are expected to yield differences in estimated valuation. However, the appropriate magnitude of adjustment is not clear, because of uncertainty about the value(s) of the income elasticity and very little empirical evidence concerning the relationship between wealth and mortality valuation.

A second potential adjustment is to convert all estimates into marginal changes in consumer income (net of taxes and benefits). In hedonic-wage studies, workers choices are in principle driven by comparing the total incremental compensation with the total incremental risk between jobs, where total compensation includes wages, health insurance, retirement income, compensation conditional on injury, and other benefits, all evaluated post tax. In stated-preference studies, respondents are likely to view payments as coming from post-tax income (in principle, respondents may be asked about payments that would be made using either pre- or post-tax income; this detail is usually not specified but may be inferred from question wording). Adjustment for these factors is difficult because of variation in marginal tax rates and benefit schedules across populations, and so the committee does not view it as critical, but suggests that research attention be directed toward determining whether such adjustments can be made.

8. *What reporting and other protocols should the researchers conducting the combination study follow? How should the analysis handle zero or negative mortality risk valuation estimates from studies that otherwise meet its selection criteria for inclusion?*

The purpose of the study should be clearly defined. For what population is the study attempting to combine estimates? There should also be an explicit description of the rules for the inclusion and exclusion of items in the combination study, and of the search rule by which the candidate studies were identified in the first place.

There should be systematic coding of important features of the items included in the combination study, for example the metric in which wages are expressed, the metric for risk itself, the other variables included in models involving VSL, and the populations for which VSL was evaluated.

Regarding the exclusion of zero of negative values, the committee notes that, due to sampling variation, it is certainly possible for studies, even those that satisfy the criteria of an acceptable study to obtain zero or negative estimates for the VSL. We therefore urge that zero or negative estimates not be excluded from a combination study.

9. *What future research or additional data would offer the most improvement in the Agency's ability to derive summary estimates of mortality risk valuation for environmental policy analyses over the short run? What longer-term research is most needed for improved summary mortality risk valuation estimates?*

- Fund more studies that will examine how the VSL varies with age and health status, the empirical correlates of life expectancy.
- Fund more studies that will shed light on the relationship between wealth and mortality valuation (income elasticity of VSL).
- Reanalyze the Pope et al. data to determine whether the impact of air pollution on mortality varies with age, rather than using a constant proportional hazard model).
- Attempt to improve hedonic wage estimates of the value of mortality risk reductions. Existing estimates, as pointed out by Dan Black and co-authors, suffer from omitted variable bias problems and problems of measurement error (in measuring risk of death), which cause estimates of the VSL to be biased.
- Consider combining RP and SP estimates using a structural approach, as in Smith, Pattanayak and Van Houtven (2006).

LIFE EXPECTANCY CHARGE QUESTIONS

Charge Questions 1 & 2:[1]
What is the most appropriate methodology to use when valuing changes in mortality risk for persons with different remaining life expectancies? Is it appropriate to use a standard VSL to value reductions in mortality risk when information on remaining life expectancy is not available?

It is anticipated that EPA will need to issue rules affecting persons who differ in their remaining life expectancies in a relatively short time-frame. What does existing research imply about approaches to valuing mortality risk when people have life expectancies of varying lengths? How applicable and relevant is the existing literature and how does the existing theoretical and empirical literature inform these issues?

According to standard welfare economics the value of a reduction in mortality risk (e.g., in the probability of dying over a stated period) is what a person is willing to pay for it. This amount may be affected by a person's remaining life expectancy, but theory (e.g., the lifecycle consumption model with uncertain lifetime) has little to say about the relationship between willingness to pay (WTP) and remaining life expectancy. Only in very special cases can it be said that WTP should be an increasing function of remaining life expectancy.[2]

The relationship between WTP for mortality risk changes and remaining life expectancy is, therefore, an empirical matter. Unfortunately, this relationship is difficult to measure since remaining life expectancy is not observable while an individual is still alive. Individuals could be asked in a stated preference study what they would pay to reduce their probability of dying, *assuming different life expectancies*. However, this is a difficult question. In revealed and stated preference studies all that can be observed *ex ante* are correlates of life expectancy: viz., age and health status. So, one could try to measure how WTP varies with age and health status.

This suggests that EPA may, in principle, want to allow WTP to vary with age and health status. It is, however, the committee's judgment that the empirical literature is not advanced enough at present to provide clear guidance as to how age and health status affect WTP for changes in mortality risk. This suggests that EPA should, at present, use an age-independent VSL to value mortality risk reductions according to the conventional paradigm. However, we also urge the Agency to report the age distribution of statistical lives saved and the average remaining life expectancies of persons in each age group.

[1] These questions have been reworded to better reflect the Agency's intent, as discussed during the September 15 meeting.
[2] For example, in a lifecycle model with uncertain lifetime and perfect annuities markets, WTP for a change in the conditional probability of dying at any age will be an increasing function of discounted remaining life expectancy if an individual consumes at a constant rate throughout his lifetime and has a period utility function that is independent of age.

Charge Question 3:
Are there other areas of the literature that should be examined and how would they inform this issue in the short tem (i.e., less than 6 months)?

The committee agrees that the White Paper by Dockins, Maguire, and Simon covered the appropriate literature.

Charge Question 4:
What type of long-term research can inform these issues?

The committee agrees that willingness to pay for risk reduction is likely to be affected by remaining life expectancy, which is related to both age and baseline health status. The existing evidence on these relationships is weak and occasionally contradictory. The committee recommends that additional research be funded to improve these estimates.

Charge Question 5:
What paradigms should be considered in valuing changes in mortality risk for person with different life expectancies? How will these paradigms inform us in the short term?

One paradigm that is commonly used to allow remaining life expectancy to affect the value of a reduction in mortality risk is the *Value of a Statistical Life Year (VSLY)*. The VSLY assumes that the value of mortality risk reductions is proportional to remaining life expectancy (or discounted remaining life expectancy) and uses this assumption to calculate a value per life year saved. The VSL is then computed by multiplying the average remaining life expectancy of the population affected by a regulation by the value per life year saved. This procedure is difficult to justify on either theoretical or empirical grounds, if the appropriate valuation concept is what a person would pay to reduce his own risk of dying. There is no empirical evidence to suggest that the VSLY is constant, or that the VSL declines in proportion to remaining life expectancy, which the constant VSLY implies. To apply the VSLY correctly would require first estimating how the VSLY varies with age. If this can be done, it would be simpler to use an age-adjusted VSL than using an age-adjusted VSL to calculate and age-adjusted VSLY.

Charge Question 6:
More generally, based on the economics literature, under what conditions is it most important to provide information on life expectancy and baseline risks as part of an economic analysis of environmental policy? If the information cannot be incorporated directly into monetized benefits estimates, how might it best be provided as a supplemental analysis?

In general, the measure of benefits based on WTP should reflect the WTP of the population that is affected by the change. The central, default VSL of $6.1 million, which is based on 21 hedonic wage and 5 stated preference estimates, reflects the WTP of the population included in these studies. The wage studies obviously reflect a population of working individuals, which implies at least some minimal health status (i.e., healthy enough to work) and a specific age distribution (i.e., working age adults) with an associated average life expectancy (estimated to be 35 years). If the population most affected by an EPA regulation or policy change differs from the population represented in these studies, then the WTP estimates generated by these studies will be a biased estimated of the true WTP of the affected population. *Whenever* this is the case, it will be important to provide information on the life expectancy and baseline risk[3] of the affected population as part of an economic analysis of the policy.

Unfortunately, the current economics literature does not provide convincing evidence regarding the direction of the bias that would exist if the baseline risk and life expectancy of the affected population differ from those of the population included in the WTP studies, i.e., there is mixed evidence on whether increases in baseline risk or reductions in life expectancy increase or decrease WTP estimates (see discussion of other charge questions). Nonetheless, it is important to include these characteristics of the affected population for two reasons: (1) to highlight the potential for bias, even if it is not possible to predict its direction, and (2) to highlight the fact that the policy is likely to affect certain sub-populations disproportionately. While this latter information may not be formally incorporated into the benefit-cost analysis (e.g., by providing WTP estimates that are specific to affected sub-populations), it would provide the basis for an equity assessment, which in many cases is required by statute, executive order, or agency policy. Information about disproportionate impacts can be important input into policy decisions.

The information about baseline risk and life expectancy is most useful if provided in the form of a distribution (rather than simply an average) across the affected population. This is particularly true when the distribution is bi-modal. A bi-modal distribution would exist, for example, in cases where the very young and the very old are susceptible to pollution effects.

[3] Strictly speaking, the term baseline risk refers to the survival curves of the members of the population affected by the regulation.