

**APPENDIX F**

**An Empirical Bayes Approach to Combining Estimates of the Value of a Statistical  
Life for Environmental Policy Analysis**

**by Ikuho Kochi, Bryan Hubbell, and Randall Kramer**

May 12, 2003

## **An Empirical Bayes Approach to Combining and Comparing Estimates of the Value of a Statistical Life for Environmental Policy Analysis\***

**Ikuho Kochi**

*Nicholas School of the Environment and  
Earth Sciences, Duke University*

**Bryan Hubbell**

*U.S. Environmental Protection Agency*

**Randall Kramer<sup>†</sup>**

*Nicholas School of the Environment and  
Earth Sciences, Duke University*

---

\* The authors thank David Higdon for comments on an earlier draft of this manuscript. Please do not cite without permission.

<sup>†</sup> Corresponding author: Randall Kramer, Nicholas School of the Environment and Earth Sciences, Duke University, Durham, North Carolina 27708-0328 USA, email: [kramer@duke.edu](mailto:kramer@duke.edu).

**Abstract**

An empirical Bayes pooling method is used to combine and compare estimates of the Value of a Statistical Life (VSL). The data come from 40 selected studies published between 1974 and 2000, containing 196 VSL estimates. The estimated composite distribution of empirical Bayes adjusted VSL has a mean of \$5.4 million and a standard deviation of \$2.4 million. The empirical Bayes method greatly reduces the variability around the pooled VSL estimate. The pooled VSL estimate is sensitive to the choice of valuation method and study location, but not to the source of data on occupational risk.

**Key words:** Value of a Statistical Life (VSL), empirical Bayes estimate, environmental policy, health policy, contingent valuation method, hedonic wage method

**JEL subject category number: J17, C11, Q28**

The value of a statistical life is one of the most controversial and important components of any analysis of the benefits of reducing environmental health risks. Health benefits of air pollution regulations are dominated by the value of premature mortality benefits. In recent analyses of air pollution regulations (United States Environmental Protection Agency (USEPA), 1999), benefits of reduced mortality risks accounted for well over 90 percent of total monetized benefits. The absolute size of mortality benefits is driven by two factors, the relatively strong concentration-response function, which leads to a large number of premature deaths predicted to be avoided per microgram of ambient air pollution reduced, and the value of a statistical life (VSL), estimated to be about \$6.3 million<sup>1</sup>. In addition to the contribution of VSL to the magnitude of benefits, the uncertainty surrounding the mean VSL estimate accounts for much of the measured uncertainty around total benefits. Thus, it is important to obtain reliable estimates of both the mean and variance of VSL.

The VSL is the measurement of the sum of society's willingness to pay (WTP) for one unit of fatal risk reduction (i.e. one statistical life). Rather than the value for any particular individual's life, the VSL represents what a whole group is willing to pay for reducing each member's risk by a small amount (Fisher et al. 1989). For example, if each of 100,000 persons is willing to pay \$10 for the reduction in risk from 2 deaths per 100,000 people to 1 death per 100,000 people, the VSL is \$1 million ( $\$10 \times 100,000$ ). Since fatal risk is not directly traded in markets, non-market valuation methods are applied to determine WTP for fatal risk reduction. The two most common methods for obtaining estimates of VSL are the revealed preference approach including hedonic wage and hedonic price analyses, and the stated preference approach including contingent valuation, contingent ranking, and conjoint methods. EPA does not conduct

original studies but relies on existing VSL studies to determine the appropriate VSL to use in its cost-benefit analyses. The primary source for VSL estimates used by EPA in recent analyses has been a study by Viscusi (1992). Based on the VSL estimates recommended in this study, EPA fit a Weibull distribution to the estimates to derive a mean VSL of \$6.3 million, with a standard deviation of \$4.2 million (U.S. EPA, 1999).

We extend Viscusi's study by surveying recent literature to account for new VSL studies published between 1992 and 2001. This is potentially important because the more recent studies show a much wider variation in VSL than the studies recommended by Viscusi (1992). The estimates of VSL reported by Viscusi range from 0.8 to 17.7 million. More recent estimates of VSL reported in the literature range from as low as \$0.1 million per life saved (Dillingham, 1985), to as high as \$87.6 million (Arabsheibani and Marin, 2000). Careful assessment is needed to determine the plausible range of VSL, taking into account these new findings.

There are several potential methods that can be used to obtain estimates of the mean and distribution of VSL. In a study prepared under section 812 of the Clean Air Act Amendments of 1990 (henceforth called the EPA 812 report), it was assumed that each study should receive equal weight, although the reported mean VSL in each study differs in its precision. For example, Leigh and Folson (1984) estimate a VSL of \$10.4 million with standard error of \$5.2 million, while Miller (1997) reports almost the same VSL (\$10.5 million) but with a much smaller standard error (\$1.5 million)<sup>2</sup>. As Marin and Psacharopoulos (1982) suggested, more weight should be given to VSL estimates that have smaller standard errors.

Our focus is to develop a more statistically robust estimate of the mean and distribution of VSL using the empirical Bayes estimation method in a two-stage pooling model. The first stage groups individual VSL estimates into homogeneous subsets to provide representative

sample VSL estimates. The second stage uses an empirical Bayes model to incorporate heterogeneity among sample VSL estimates. This approach allows the overall mean and variance of VSL to reflect the underlying variability of the individual VSL estimates, as well as the observed variability between VSL estimates from different studies. Our overall findings suggest the empirical Bayes method provides a pooled estimate of the mean VSL with greatly reduced variability. In addition, we conduct sensitivity analyses to examine how the pooled VSL is affected by valuation method, study location, source of occupational risk data and the addition of estimates with missing information on standard errors. This sensitivity analysis allows us to systematically compare VSL estimates to determine how they are influenced by study design characteristics.

## **1. Methodology**

### **1.1 Study selection**

We obtained published and unpublished VSL studies by examining previously published meta-analysis or review articles, citations from VSL studies and by using web searches and personal contacts.

The data were prepared as follows. First, we selected qualified studies based on a set of selection criteria applied in Viscusi (1992). Second, we computed and recorded all possible VSL estimates and associated standard errors in each study. Third, we made subsets of homogeneous VSL estimates and calculated the representative VSL for each subset by averaging VSLs and their standard errors<sup>3</sup>. Each step is discussed in detail below.

Since the empirical Bayes estimation method (pooled estimate model) does not control for the overall quality of the underlying studies, careful examination of the studies is required for

selection purposes. In order to facilitate comparisons with the EPA 812 report, we applied the same selection criteria that were applied in that report, based largely on the criteria proposed in Viscusi (1992).

Viscusi (1992) examined 37 hedonic wage (HW), hedonic price (HP) and contingent valuation (CV) studies of the value of a statistical life, and listed four criteria for determining the value of life for policy applications. The first criterion is the choice of risk valuation method. Viscusi (1992) found that all the HP studies evaluated failed to provide an unbiased estimate of the dollar side of the risk-dollar tradeoff, and tend to underestimate VSL. Therefore only HW studies and CV studies are included in this study.

The second criterion is the choice of the risk data source for HW studies. Viscusi argues that actuarial data reflect risks other than those on the job, which would not be compensated through the wage mechanism, and tend to bias VSL downward. Therefore some of the initial HW studies that used actuarial data are removed from this analysis. The third criterion is the model specification in HW studies. Most studies apply a simple regression of the natural log of wage rates on risk levels. However, a few of the studies estimate the tradeoff for discounted expected life years lost rather than simply risk of death. This estimation procedure is quite complicated, and the VSL estimates tend to be less robust than in a simple regression estimation approach. Only studies using the simple regression approach are used in this analysis.

The fourth criterion is the sample size for CV studies. Viscusi argues that the two studies he considered whose sample sizes were 30 and 36 respectively were less reliable and should not be used. In this study, a threshold of 100 observations was used as a minimum sample size<sup>4</sup>.

There are several other selection criteria that are implicit in the 1992 Viscusi analysis<sup>5</sup>. The first is based on sample characteristics. In the case of HW studies, he only considered studies that examined the wage-risk tradeoff among general or blue-collar workers. Some recent studies only consider samples from extremely dangerous jobs, such as police officer. Workers in these jobs may have different risk preferences and face risks much higher than those evaluated in typical environmental policy contexts. As such, we exclude those studies to prevent likely downward bias in VSL relative to the general population. In the case of CV studies, Viscusi only considered studies that used a general population sample. Therefore we also exclude CV studies that use a specific subpopulation or convenience sample, such as college students.

The second implicit criterion is based on the location of the study. Viscusi (1992) considered only studies conducted in high income countries such as U.S., U.K. and Japan. Although there are increasing numbers of CV or HW studies in developing countries such as Taiwan, Korea and India, we exclude these from our analysis due to differences between these countries and the U.S. Miller (2000) found that income level has a significant impact on VSL, and because we are seeking a VSL applicable to U.S. policy analysis, inclusion of VSL estimates from low-income countries may bias VSL downward. In addition, there are potentially significant differences in labor markets, health care systems, life expectancy, and preferences for risk reductions between developed and developing countries. Thus, our analysis only includes studies in high-income OECD member countries<sup>6</sup>. Finally, our analysis only uses studies that estimate people's WTP for immediate risk reduction due to concerns about comparisons between risks with long latency periods with inherent discounting or uncertainty about future baseline health status.

## 1.2 Data preparation

In VSL studies, authors usually report the results of a hedonic wage regression analysis, or WTP estimates derived from a CV survey. In the studies we reviewed, a few authors reported all of the VSL that could be estimated based on their analysis, but most authors reported only selected VSL estimates and provided recommended VSL estimates based on their professional judgment. This judgment subjectively takes into account the quality of analysis, such as the statistical significance of the result, the target policy to be evaluated, or judgments based on comparative findings. Changes in statistical methods and best practices for study design during the period covered by our analysis may invalidate the subjective judgments used by authors to recommend a specific VSL. To minimize potential judgment biases, as well as make use of all available information, we re-estimate all possible VSLs based on the information provided in each study and included them in our analysis as long as they met the basic criteria laid out by Viscusi (1992)<sup>7</sup>. For certain specifications some authors found a negative VSL. However, in every case the authors rejected the plausibility of the negative estimates. We agree that negative VSL are highly implausible and exclude them from our primary data set. However, we do present sensitivity analysis showing the effects of excluding the negative estimates.

### *Estimation of VSL from HW studies*

Most of the selected HW studies use the following equation to estimate the wage-risk premium:

$$\ln Y_i = a_1 p_i + a_2 q_i + a_3 p_i^2 + X_i \beta + \varepsilon_i \quad (1)$$

where  $Y_i$  is equal to earnings of individual  $i$ ,  $p_i$  and  $q_i$  are job related fatal and non-fatal risk faced by  $i$  ( $q_i$  often omitted),  $X_i$  is a vector of other relevant individual and job characteristics (plus a

constant) and  $\varepsilon_i$  is an error term. In many cases, the wage equation will also include fatal risk squared and interactions between risk and variables such as union status. Based on equation (1), the VSL is estimated as follows.

$$VSL = (d\ln Y / dp_i) \times \text{mean annual wage}^8 \times \text{unit of fatal risk}^9 \quad (2)$$

Note that  $d\ln Y / dp_i$  may include terms other than  $a_1$  if there are squared or interaction terms.

VSL is usually evaluated at the mean annual wage of the sample population. The unit of fatal risk is the denominator of the risk statistic, i.e. 1000 if the reported worker's fatal risk is 0.02 per 1000 workers. If there is an interaction term between fatal risk and human capital variables such as "Fatal Risk"  $\times$  "Union Status", the VSL is evaluated at the mean values of the union status variable. If there is a squared risk term, the VSL is evaluated at the mean value of fatal risk.

*Estimation of standard error of VSL from HW studies*

The standard error of the VSL (SE(VSL)) from a HW study is  $Var(VSL) = (\text{unit of risk})^2 Var(\partial \ln Y / \partial p \times \bar{Y})$ , where  $\bar{Y}$  is the average wage for the sample.

For example, if the wage equation is specified as  $\ln Y = a_1 p_i + a_2 q_i + a_3 p_i^2 + a_4 p_i UNION + \varepsilon_i$ , then

$$Var(VSL) = (\text{unit of risk})^2 [Var(a_1 \bar{Y}) + 4Var(a_2 \bar{p} \bar{Y}) + Var(a_3 \bar{Y} \overline{UNION}) + 2Cov(a_1 \bar{Y}, a_2 \bar{p} \bar{Y}, a_3 \bar{Y} \overline{UNION})]$$

To calculate the full variance, allowing for the observed variability in wages and fatal risk, one needs to calculate the variance of the product of the regression coefficients and the wage, risk, and interaction terms. We use the formula for the exact variance of products provided by Goodman (1960). For the first variance term above, this formula would be

$$Var(a_1 \bar{Y}) = \bar{Y}^2 \frac{s^2(a_1)}{n} + a_1^2 \frac{s^2(\bar{Y})}{n} - \frac{s^2(a_1) s^2(\bar{Y})}{n^2}$$

Most of the studies included in our analysis do not report the variance of annual wage or the covariance matrix (either for the parameter estimates or the variables), so we calculated the standard error of VSL based on the available information, usually consisting of the standard errors of the estimated parameters of the wage equation. In this case the variance formula reduces to

$$Var(VSL) = (\text{unit of risk})^2 \left[ \bar{Y}^2 Var(a_1) + 4\bar{p}^2 \bar{Y}^2 Var(a_2) + \bar{Y}^2 \overline{UNION}^2 Var(a_3) \right]$$

To assess the impact of treating mean annual wage as a constant, we estimate the standard error with and without the wage variance for the 45 VSL estimates for which information on the variance of wage was available. We find that the differences between the two estimates of standard error are fairly small, within \$0.2 million for most estimates. In no case does the standard error differ by more than 10 percent. We also assess the impact of omitting the covariance term by comparing the reported standard error of Scotton and Taylor (2000) providing a “full” variance estimate for the estimated VSL with our estimated standard error, which does not include the covariance term. We find that the difference in standard error is quite small. Note that the published standard error from this study treats mean annual wage as fixed, so the comparison shows only the effect of excluding the covariance term. These results suggest the impact of omitting the covariance terms and treating mean annual wage as fixed in our calculation of standard errors should not have a significant effect on our results.

#### *Estimation of VSL and standard error from CV studies*

For most of the CV surveys, we could not estimate the VSL and its standard error unless the author provided mean or median WTP and a standard error for a certain amount of risk reduction. When this information is available, the VSL and its standard error are simply

calculated as WTP divided by the amount of risk reduction, and SE(WTP) divided by the amount of risk reduction, respectively.

*Estimation of representative VSL for each study*

Most studies reported multiple VSL estimates. For the empirical Bayes approach, which we use in our analysis, each estimate is assumed to be an independent sample, taken from a random distribution of the conceivable population of studies. This assumption is difficult to support given the fact that there are often multiple observations from a single study. To solve this problem, we constructed a set of homogeneous (and more likely independent) VSL estimates by employing the following approach.

We arrayed individual VSL estimates by study author (to account for the fact that some authors published multiple articles using the same underlying data). We then examined homogeneity among sub-samples of VSL estimates for each author by using Cochran's Q-statistics. The test statistic Q is the sum of squares of the effect about the mean where the  $i_{th}$  square is weighted by the reciprocal of the estimated variance. Under the null hypothesis of homogeneity, Q is approximately a  $\chi^2$  statistic with  $n - 1$  degrees of freedom (DerSimonian and Laird, 1986). If the null hypothesis was not rejected, we take the average of the VSL for the subset and the standard error to estimate the representative mean VSL for that author.

If the hypothesis of homogeneity was rejected, we further divided the samples into subsets according to their different characteristics such as source of risk data and type of population (i.e. white collar or blue collar), and tested for homogeneity again. We repeated this process until all subsets were determined to be homogeneous.

### 1.3 The empirical Bayes estimation model

In general, the empirical Bayes estimation technique is a method that adjusts the estimates of study-specific coefficients ( $\beta$ 's) and their standard errors by combining the information from a given study with information from all the other studies to improve each of the study-specific estimates. Under the assumption that the true  $\beta$ 's in the various studies are all drawn from the same distribution of  $\beta$ 's, an estimator of  $\beta$  for a given study that uses information from all study estimates is generally better (has smaller mean squared error) than an estimator that uses information from only the given study (Post et al. 2001).

The empirical Bayes model assumes that

$$\beta_i = \mu_i + e_i \quad (6)$$

where  $\beta_i$  is the reported VSL estimate from study  $i$ ,  $\mu_i$  is the true VSL,  $e_i$  is the sampling error and  $N(0, s_i^2)$  for all  $i = 1, \dots, n$ . The model also assumes that

$$\mu_i = \mu + \delta_i \quad (7)$$

where  $\mu$  is the mean population VSL estimate,  $\delta_i$  captures the between study variability, and  $N(0, \tau^2)$ ,  $\tau^2$  represents both the degree to which effects vary across the study and the degree to which individual studies give biased assessments of the effects (Levy et al., 2000; DerSimonian and Laird, 1986).

The weighted average of the reported  $\beta_i$  is described as  $\mu_w$ . The weight is a function of both the sampling error ( $s_i^2$ ) and the estimate of the variance of the underlying distribution of  $\beta$ 's ( $\tau^2$ ). These are expressed as follows;

$$\mu_w = \frac{\sum w_i^* \beta_i}{\sum w_i^*} \quad (8)$$

$$s.e. (\mu_w) = (\sum w_i^*)^{-1/2} \quad (9)$$

where  $w_i^* = \frac{1}{w_i^{-1} + \tau^2}$  and  $w_i = \frac{1}{s_i^2}$

$\tau^2$  can be estimated as

$$\tau^2 = \max \left( 0, \left( \frac{(Q - (n-1))}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}} \right) \right) \quad (10)$$

where  $Q = \sum w_i (\beta_i - \beta^*)^2$  (Cochran's Q-statistic) and  $\beta^* = \frac{\sum w_i \beta_i}{\sum w_i}$

The adjusted estimate of the  $\beta_i$  is estimated as

$$\text{Adjusted } \beta_i = \frac{\frac{\beta_i + \mu_w}{e_i + \tau^2}}{\frac{1}{e_i} + \frac{1}{\tau^2}} \quad (11)$$

This adjustment, as illustrated in Figure 1, pulls the reported estimates of  $\beta_i$  towards the pooled estimate. The more within-study variability, the less weight the  $\beta_i$  receives relative to the pooled estimate, and the more it gets adjusted towards the pooled estimate. The adjustment also reduces the variance surrounding the  $\beta_i$  by incorporating information from all  $\beta$ 's into the estimate of  $\beta_i$ . (Post et al. 2001). In our analysis,  $\beta_i$  corresponds to the VSL of the  $i$ th study.

In order to visually compare the distributions, we used kernel density estimation to develop smooth distributions based on the empirical Bayes estimate. The kernel estimation provides a smoother distribution than the histogram approach. The kernel estimator is defined

by  $f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$ . The kernel function,  $\int_{-\infty}^{\infty} K(x)dx = 1$ , is usually a symmetric

probability density function, e.g. the normal density, and  $h$  is window width. The kernel function  $K$  determines the shape of the bumps, while  $h$  determines their width. The kernel estimator is a

*May 12, 2003*

sum of ‘bumps’ placed at observations and the estimate  $f$  is constructed by adding up the bumps (Silverman 1986). We assumed a normal distribution for  $K$  and a window width  $h$  equal to 0.7, which was wide enough to give a reasonably smooth composite distribution while still preserving the features of the distribution (e.g. bumps). The choice of window width is arbitrary, but has no impact on the statistical comparison, which is described below.

To compare the different distributions of VSL, we applied the bootstrap method, which is a nonparametric method for estimating the distribution of statistics. Bootstrapping is equivalent to random sampling with replacement. The infinite population that consists of the  $n$  observed sample values, each with probability  $1/n$ , is used to model the unknown real population (Manly 1997). We first conducted re-sampling 1000 times, and compared the distributions in terms of mean, median and interquartile range.

## 2. Results and sensitivity analyses

In total, we collected 47 HW studies and 29 CV studies. A data summary for each stage of analysis is shown in Table 1. After applying the selection criteria outlined in section 2.1, there were 31 HW studies and 14 CV studies left for the analysis. In our final list, there are 22 new studies published between 1990 and 2000. We re-estimated all possible VSL for the selected studies, and obtained 232 VSL estimates.<sup>10 11</sup> There were 23 VSL estimates from five studies for which standard errors were not available, and thus they are excluded from our primary analysis, although we examine the impact of excluding those studies in a sensitivity analysis. After testing for homogeneity among sub-samples, we obtained 60 VSL subsets, and estimated a representative VSL and standard error for each subset. Finally, we applied the empirical Bayes method and obtained an adjusted VSL value for each subset.

It is worthwhile to note how the empirical Bayes approach reduces the unexplained variability among VSL estimates. Our 196 VSL estimates show an extremely wide range from \$0.1 million to \$95.5 million with a coefficient of variation of 1.3 (in 2000 constant dollars). The VSL estimates from the 60 subsets range from \$0.3 million to \$43.1 million with a coefficient of variation of 1.2, and the adjusted VSL estimates range from \$0.7 million to \$13.9 million with a coefficient of variation of 0.4.

### 2.1 The distribution of VSL

Figure 2 shows the kernel density estimates of the composite distribution of the empirical Bayes adjusted VSL (using the 60 representative VSL estimates) and the Weibull distribution for the 26 VSL estimates as reported in the EPA 812 report. The summary results are shown in Table 2. The composite distribution of adjusted VSL has a mean of \$5.4 million

with a standard error of \$2.4 million. This mean value is smaller than that based on the EPA 812 Weibull distribution and has less variance (EPA 812's coefficient of variation is 0.7) even though our VSL sample has a range more than five times as wide as the EPA 812 sample.

## **2.2 Sensitivity analyses**

### **2.2.1 Sensitivity to choice of valuation method**

Many researchers argue that the VSL is sensitive to underlying study characteristics (Viscusi 1992, Carson, et al. 2000, Mrozek and Taylor 2002). One of the most interesting differences is in the choice of valuation method. To determine if there is a significant difference between the empirical Bayes adjusted distributions of VSL using HW and CV estimates, we used bootstrapping to test the hypothesis that HW and CV estimates of VSL are from the same underlying distribution.

We divided the set of VSL studies into HW and CV and applied the homogeneity subsetting process and empirical Bayes adjustment method to each group. The kernel density estimates of the distributions for HW and CV sample are shown in Figure 3. The HW distribution has a mean value of \$9.4 million with a standard error of \$4.7 million while the CV distribution has much smaller mean value of \$2.8 million with a standard error of \$1.3 million (see Table 2). Bootstrap tests of significance show the VSL based on HW is significantly larger than that of CV ( $p < 0.001$ ), comparing means, medians and interquartile ranges between the distributions.

### **2.2.2 Sensitivity to study location**

Because of differences in labor markets, health care systems, and societal attitudes towards risk, VSL estimates from HW studies may potentially be sensitive to the country in which the study was conducted (this may also be true for CV studies, however there were too few CV estimates to conduct similar comparisons). Empirical Bayes estimation was applied to HW samples from the U.S. and U.K. separately. (Comparisons with Canada and Australia were not conducted because of small sample sizes for those countries.) The distribution for the U.S. sample has a mean value of \$8.5 million with a standard error of \$4.9 million, while the distribution for the U.K. sample has a mean value of \$22.6 million with a standard error of \$4.9 million. Bootstrap tests of significance show that the U.S. estimates are significantly different from UK estimates based on comparing means and medians between distributions.

### **2.2.3 Sensitivity to source of occupational risk data**

Moore and Viscusi (1988) found that VSL was sensitive to choice of source of occupational risk data. According to their results, the VSL estimated based on Bureau of Labor Statistics (BLS) death-risk data is significantly smaller than that estimated based on National Institute of Occupational Safety and Health (NIOSH) death risk data. We estimated the empirical Bayes adjusted VSL distribution for each risk data source, and we did not find a significant difference between the two distributions. However, the reliability of our result is limited due to the small number of studies based on the BLS risk data.

#### **2.2.4 Sensitivity to excluded VSL estimates**

We also examined the sensitivity of our results to excluded estimates. To do this, we added to the sample the VSL estimates that were excluded from the primary analysis due to the lack of a standard error. We assumed for this test that all reported VSL estimates should have passed at least a 95 percent significance test, and estimate the corresponding standard error at this significance level for each VSL. This added nine averaged VSL estimates to the set of 60 representative estimates, including four estimates from HW studies and five from CV studies.

The distribution of the enhanced sample has a mean value of \$4.7 million with a standard error of \$2.2 million. Compared with the result of our main analysis, the mean value is reduced by \$0.7 million. This is because we have added more estimates from CV, which tends to produce relatively lower VSL. Bootstrap tests of significance show the VSL from HW studies is still significantly different from that from CV studies ( $p < 0.0001$ ), comparing means, medians and interquartile ranges.

We also report a 5% trimmed mean that increases the combined mean from both valuation methods from \$5.4 million to \$5.8 million with no effect on the coefficient of variation. Finally, we consider the impact of including negative estimates. Since these estimates were all associated with HW studies, the HW mean drops from \$9.4 million to \$6.6 million. This also has a noticeable effect on the combined mean dropping it from \$5.4 million to \$4.1 million. The difference between the CV and HW estimates remains significant based on bootstrap tests of the means and medians.

### **3. Conclusions**

The meta analysis we have used results in a composite distribution of empirical Bayes adjusted VSL with a mean of \$5.4 million and a standard deviation of \$2.4 million. This is a somewhat lower mean than previous pooled estimates, and because of the Bayesian adjustment process, there is greatly reduced variability as evidenced by the coefficient of variation even though our dataset has a much wider range than previous studies.

Starting from a baseline of the literature used in Viscusi (1992), our approach has generated a set of hypotheses that may challenge some previously held assumptions. It is clear that VSL analysts need to look closely at study location; our estimates show significant differences in VSL even between developed countries with relatively similar income levels. It is also important to look at valuation method as we found quite different VSL estimates in the hedonic wage versus contingent valuation datasets. Our finding that the hedonic method generates significantly larger estimates than the CV approach is consistent with a comparison of CV and revealed preference approaches to valuing quasi-public goods reported by Carson (1996).

Theoretically, the two valuation methods should not necessarily provide the same results because the HW approach is estimating a local trade-off, while the CV approach approximates a movement along a constant expected utility locus (Viscusi and Evans 1990, Lanoie, Pedro and Latour 1995). However, the impact and direction of this difference had not been systematically investigated prior to this analysis

Our sensitivity analysis found no significant difference on average in the VSL estimates between studies using BLS or NIOSH data. Additional research into appropriate measures of risk is needed. Recent work by Black (2001) suggests that measurement errors in

estimates of fatal risk can lead to large downward biases in estimates of VSL.

Aggregate level comparisons as we have done in this paper are useful in comparing the overall distribution of VSL estimates from each method, however the resulting comparisons might be significantly affected by differences in the design of each study, as the large variance in the HW distribution suggests. This problem could be addressed by applying meta-regression analysis, which can determine the impact of specific study factors by taking into consideration study characteristics such as sample population, study location, or sources of risk data (Levy et al., 2000; Mrozek and Taylor, 2002; Viscusi and Aldy, 2002).

Study location does seem to matter, but additional investigation is necessary to identify why there are differences. Simply lumping countries together as developed or developing may not be the best way to account for potential differences in VSL. Differences in health care system may be a potential factor, as there are a number of differences in insurance coverage and access to health care across developed countries (Anderson and Hussey, 2000). There may be numerous other socio-cultural factors that can cause VSL estimates to diverge.

As the excluded studies sensitivity analysis indicates, our results are sensitive to the addition of small magnitude VSL estimates with low variances. For example, Krupnick et al. (2000) estimated the VSL as \$1.1 million with a standard error of \$0.05 million. If we remove this estimate from our main analysis, the overall mean VSL is increased to \$5.9 million, implying that one study reduces the overall mean by \$0.5 million. It is thus especially important to determine the reliability of CV studies very carefully by assessing any potential questionnaire and scope effects (Hammitt and Graham, 1999). Also, it may be important to investigate why the VSL estimates from CV studies are so similar despite the differences in type of risk, study location and survey method.

*May 12, 2003*

In addition to the application of the empirical Bayes method, our analysis demonstrates the importance of adopting a two-stage procedure for combining evidence from the literature when multiple estimates are available from a single source of data. The first stage sorting process using the Cochran's Q test for homogeneity seems a reasonable approach to control for over-representation of any one dataset. From the original set of 40 studies, we obtained 196 VSL estimates and then classified these into 60 homogeneous subsets. This suggests that there was a high probability of assigning too much weight to some estimates if a single stage process were used, treating each of the 196 estimates as independent. Also, the two-stage approach does not discard information from each study. Instead it uses all the available information in an appropriate manner.

As in the field of epidemiology, the economics profession should consider developing protocols for combining estimates from different studies for policy purposes. Consistent reporting of both point estimates of VSL and standard errors, or variance-covariance matrices would enhance the ability of future researchers to make use of all information in constructing estimates of VSL for policy analysis. Additional research is needed to understand how VSL varies systematically with underlying study attributes, such as estimation method or location of studies. The empirical Bayes approach outlined here provides a useful starting point in developing the variables needed for such studies.

The widely cited pooled estimate of \$6.3 million from the EPA 812 study based on Viscusi's assessment of the VSL literature was derived from a simple histogram method. This early approach ignored within and between study variability. Mrozek and Taylor presented an alternative method for deriving a mean VSL estimate for policy purposes based on a best fit regression model using only the hedonic wage studies. . We examine both CV and HW studies

*May 12, 2003*

and present a different methodology using all available information to adjust individual VSL estimates based on the within and between study variability. By generating distributions of VSL, the method allows us to test individual hypotheses regarding study attributes. These comparisons have generated a number of hypotheses that should form the foundation for future meta-analyses of VSL combining the CV and HW approaches.

**References:**

- Anderson, G.F. and P.S. Hussey, *Multinational Comparison of Health Systems Data 2050*, The Commonwealth Fund, International Health Policy Report (October 2000).
- Black, Dan A. "Some Problems in the Identification of the Price of Risk," Paper presented at USEPA Workshop, Economic Valuation of Mortality Risk Reduction: Assessing the State of the Art for Policy Applications, Silver Spring, Maryland, November 6-7, 2001.
- Arabsheibani, R. G., and A. Marin, "Stability of Estimates of the Compensation for Damage," *Journal of Risk and Uncertainty* 20:3 (2000), 247-269.
- Carson, Richard T., Nicholas E. Flores, and Norman F. Meade, "Contingent Valuation: Controversies and Evidence," *Environment and Resource Economics* 19:2 (2001), 173-210.
- Carson, Richard T., "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods," *Land Economics* 72:1 (1996), 80-99.
- DerSimonian, Rebecca, and Nan Laird, "Meta-analysis in Clinical Trials," *Controlled Clinical Trials* 7:3 (1986), 177-188.
- Dillingham, Alan E, "The Influence of Risk Variable Definition on Value of Life Estimates," *Economic Inquiry* 24 (April, 1985), 277-294.
- Eom, Young Sook, "Pesticide Residue Risk and Food Safety Valuation: A Random Utility Approach," *American Journal of Agricultural Economics* 76 (Nov, 1994), 760-771.
- Fisher, Ann, Lauraine G. Chestnut and Daniel M. Violette, "The Value of Reducing Risks of Death: A Note on New Evidence," *Journal of Policy Analysis and Management* 8:1 (1989), 88-100.

- Goodman, Leo A., "On the Exact Variance of Products," *American Statistical Association Journal* (Dec, 1960), 708- 713.
- Hammit, James K. and John D. Graham, "Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability?" *Journal of Risk and Uncertainty* 8:1 (1999), 33-62.
- Kneisner, Tomas J. and John D. Leeth, "Compensating Wage Differentials for Fatal Injury Risk in Australia, Japan, and the United States," *Journal of Risk and Uncertainty* 4:1 (1991), 75-90.
- Krupnick, Alan, Anna Alberini, Maureen Cropper, Nathalie Simon, Bernie O'Brien, Ron Goeree, and Martin Heintzelman, "Age, Health, and the Willingness to Pay for Mortality Risk Reductions: a Contingent Valuation Survey of Ontario Residents," Resources For the Future discussion paper 00-37 (Sep, 2000).
- Lanoie, Paul, Carmen Pedro, and Robert Latour, "The Value of a Statistical Life: A Comparison of Two Approaches," *Journal of Risk and Uncertainty* 10:3 (1995), 235-257.
- Leigh, Paul J. and Roger N. Folsom, (1984). "Estimates of the Value of Accident Avoidance at the Job Depend on Concavity of the Equalizing Differences Curve," *The Quarterly Review of Economics and Business* 24:1 (1984), 55-56.
- Levy, Jonathan I., James K. Hammit, and John D. Spengler, "Estimating the Mortality Impacts of Particulate Matter: What Can Be Learned from Between-Study Variability?" *Environmental Health Perspectives* 108:2 (2000), 108-117.
- Manly, Bryan F.J, *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2<sup>nd</sup> edition (London: Chapman & Hall, 1997).

May 12, 2003

- Marin, Alan, and George Psacharopoulos, "The Reward for Risk in the Labor Market: Evidence from the United Kingdom and Reconciliation with Other Studies", *Journal of Political Economy* 90:4 (1982), 827-853.
- Miller, Ted R., "Variations Between Countries in Values of Statistical Life," *Journal of Transport Economics and Policy* 34:2 (2000), 169-188.
- Miller, Paul, Charles Mulvey, and Keith Norris, "Compensating Differentials for Risk of Death in Australia", *Economic Record* 73:223 (1997), 363-372.
- Moore, Michael J. and W. Kip Viscusi, "Doubling the Estimated Value of Life: Results Using New Occupational Fatality Data", *Journal of Policy Analysis and Management* 7:3 (1998), 476-490.
- Mrozek, Janusz R. and Laura O. Taylor, "What Determines the Value of Life? A Meta-Analysis," *Journal of Policy Analysis and Management* (2002) (Forthcoming).
- Post, E., Hoaglin D., Deck L. and Larntz K., "An Empirical Bayes Approach to Estimating the Relation of Mortality to Exposure to Particulate Matter," *Risk Analysis* 21:5 (2001), 837-842.
- Silverman, Bernard W., *Density Estimation for Statistics and Data Analysis*, (London: Chapman and Hall, 1986).
- Smith, V. Kerry and Carol C.S. Gilbert, "The Implicit Risks to Life: A Comparative Analysis," *Economic Letters* 16: 3-4 (1984), 393-399.
- United States Environmental Protection Agency (USEPA), *The Benefits and Cost of the Clean Air Act, 1990 to 2010* (Washington: USEPA, 1999).
- Viscusi, W. Kip, *Fatal Tradeoffs: Public and Private Responsibilities for Risk* (New York: Oxford University Press, 1992).

*May 12, 2003*

Viscusi, W. Kip and Joseph E. Aldy, "The Value of Statistical Life: A Critical Review of Market Estimates throughout the World," Discussion Paper No. 392., John M. Olin Center for Law, Economics and Business, Harvard Law School, Cambridge, Massachusetts.

Viscusi, W. Kip and William N. Evans, "Utility Functions That Depend on Health Status: Estimates and Economic Implications," *American Economic Review* 80:3 (1990), 353-374.

Table 1. VSL Data Summary

	<b>HW</b>	<b>CV</b>	<b>Total</b>
<b>Number of collected studies</b>	47	29	76
<b>Number of selected studies</b>	31	14	45
<b>Number of estimated VSL</b>	181	51	232
<b>Number of positive VSL with imputed SE</b>	161	35	196
<b>Mean (million \$)</b>	12.3	3.8	10.8
<b>(Coefficient of variation)</b>	(1.2)	(1.5)	(1.3)
<b>Number of VSL subsets at 1<sup>st</sup> stage</b>	43	17	60
<b>Mean (million \$)</b>	12.4	3.8	9.8
<b>(Coefficient of variation)</b>	(1.1)	(0.8)	(1.2)
<b>Number of VSL subsets at 2<sup>nd</sup> stage</b>	43	17	60
<b>Mean (million \$)</b>	9.4	2.8	5.4
<b>(Coefficient of variation)</b>	(0.5)	(0.5)	(0.4)

Table 2. Results of Empirical Bayes Estimates and Bootstrap Tests for Distribution Comparisons (2000 dollars)

	Mean (million \$)	SD (million \$)	Coefficient of variation	Bootstrap Test		
				Mean	Median	Interquartile
<b>Distribution Comparison by Evaluation Method</b>						
Total (60)	5.4	2.4	0.4	P-value (Ho: HW = CV)		
CV (18)	2.8	1.3	0.5	<0.001	<0.001	<0.008
HW (42)	9.4	4.7	0.5			
<b>Distribution Comparison by Study Location (HW only)</b>						
USA (30)	8.5	4.9	0.6	P-value (Ho: US =UK)		
UK (7)	22.6	4.9	0.2	<0.001	<0.001	<0.403
<b>Distribution Comparison by Occupational Risk Data Source (HW only)</b>						
BLS (3)	10.3	4.3	0.4	P-value (Ho: BLS = NIOSH)		
NIOSH (21)	7.2	3.9	0.5	<0.694	<0.798	<0.734
<b>Distribution Comparison by Evaluation Method After Adding Excluded Estimates</b>						
Total	4.7	2.2	0.5	P-value (Ho: HW = CV)		
CV	2.6	1.3	0.5	<0.001	<0.001	<0.009
HW	8.7	4.6	0.5			
<b>5% trimmed estimate</b>						
Total	5.8	2.5	0.4			
<b>Including negative estimates</b>						
Total (67)	4.1	1.7	0.4	P-value (Ho: HW = CV)		
CV (18)	2.8	1.3	0.5	<.001	<.004	<.108
HW (49)	6.6	3.6	0.5			

Figure1. Illustration of Empirical Bayes Pooling

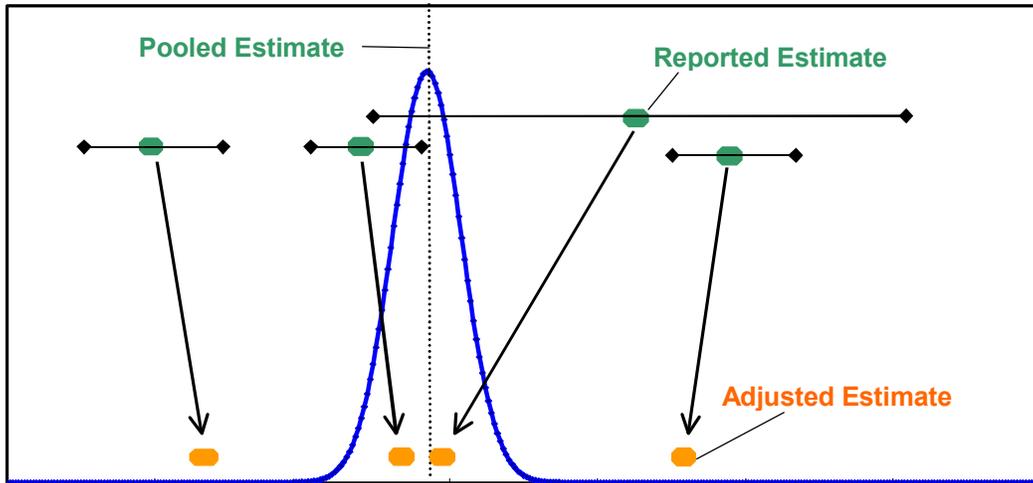


Figure 2. Comparison of Kernel Distribution of Empirical Bayes Adjusted VSL with Distribution of VSL Based on EPA Section 812 Report Estimates

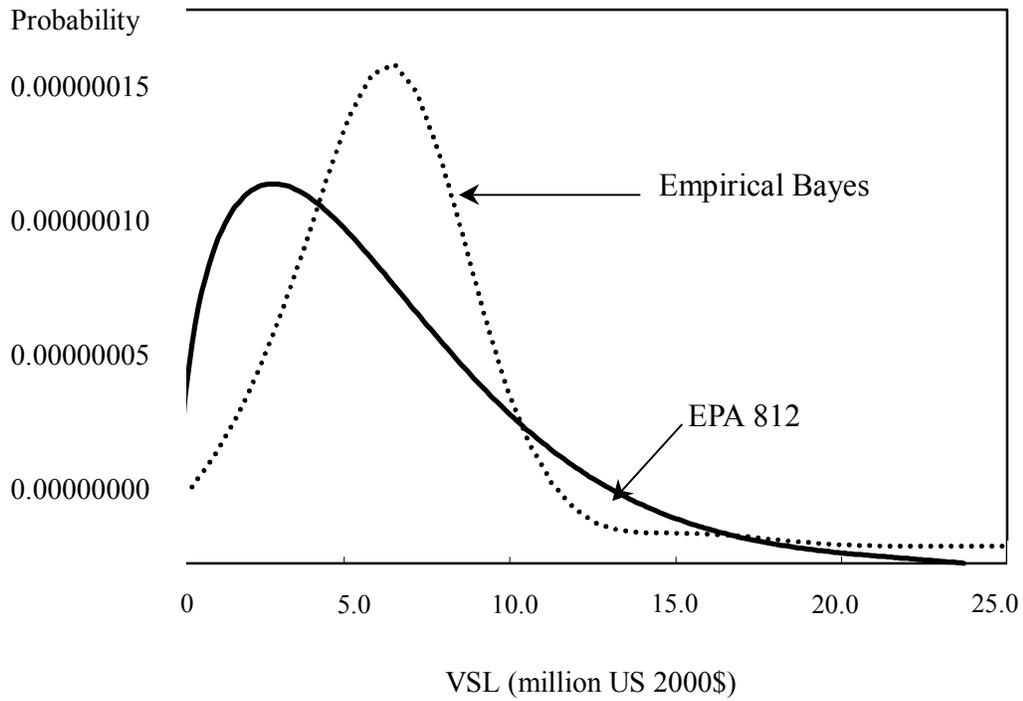
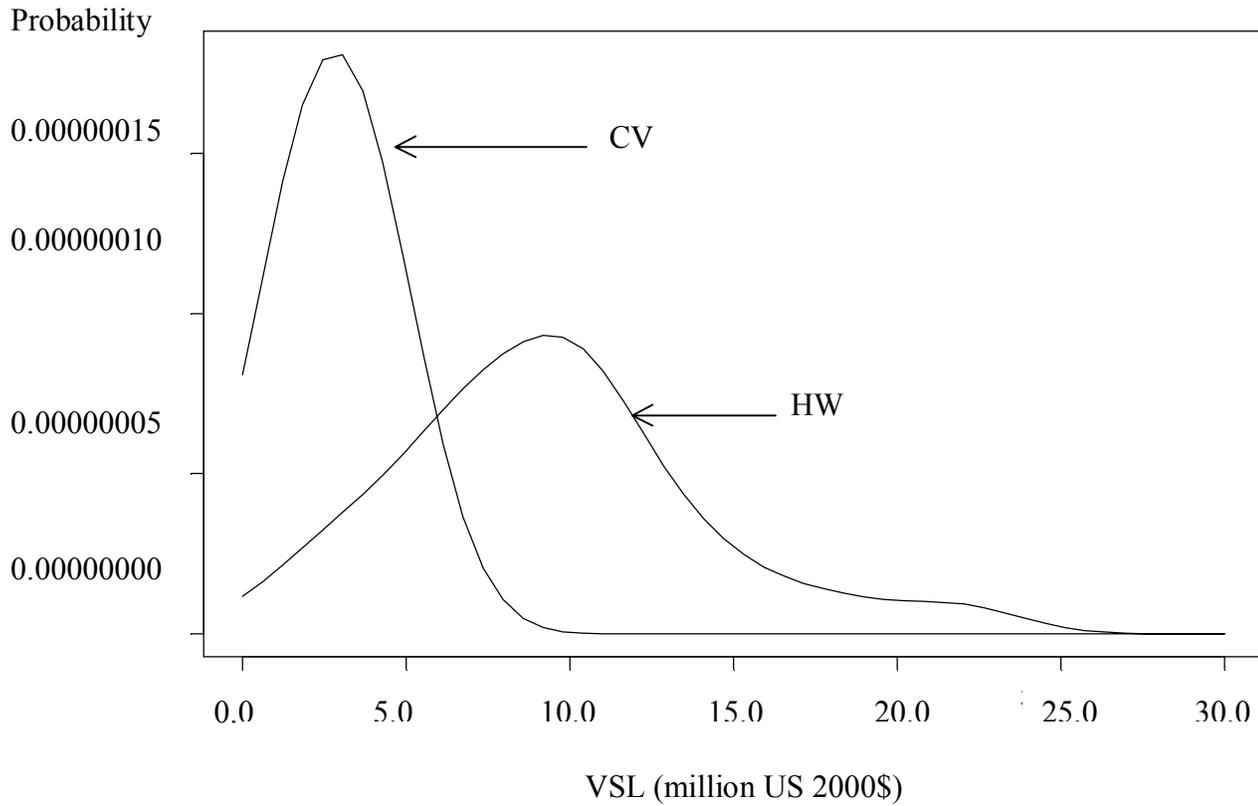


Figure 3. Comparison of Kernel Distribution of Empirical Bayes Adjusted VSL Based on HW and CV Estimates



**Notes:**

---

<sup>1</sup>All estimates reported in this paper have been converted to constant 2000 dollars using the Bureau of Labor Statistics Consumer Price Index (CPI). The CPI inflation calculator uses the average Consumer Price Index for a given calendar year. These data represent changes in prices of all goods and services purchased for consumption by urban households. For estimates reported in foreign currency, we first converted to U.S. dollars using data on Purchasing Power Parity from the Organization for Economic Cooperation and Development, and then converted to 2000 U.S. dollars using the CPI.

<sup>2</sup> Most authors do not report standard errors of VSL estimates. We have estimated the standard errors for these and other studies using an approach discussed later in the paper.

<sup>3</sup> We also employed fixed approaches for pooling, but found this resulted in an artifact of providing greater weight to studies whose authors reported multiple estimates.

<sup>4</sup> This is admittedly an arbitrary cutoff. However, we determined that a sample size of 100 did not result in many studies being excluded and smaller samples did not seem to be reasonable.

<sup>5</sup> We exclude one additional study, by Eom (1994), due to concerns about the payment context for the willingness to pay question. In that study, individuals were asked to choose between produce with different levels of price and pesticide risk. The range of potential WTP was limited by the base price of produce. In order to realize an implied VSL within the range considered by Viscusi, individuals would need to have a WTP of around \$400 per year. Because WTP in the study was tied to increases in produce prices, which ranged \$0.39 to \$1.49, it would be very unlikely that individuals would be willing to pay over a 100 times their normal price for produce to obtain the specified risk reduction. Tying WTP to observed prices thus limits the usefulness of this study for benefits transfer.

---

<sup>6</sup> From <http://worldbank.org/data/databytopic/class.htm>. High-income OECD member have annual income greater than \$9,266 per capita.

<sup>7</sup> One reviewer suggested that some published VSL estimates should be excluded from our analysis because the authors judged these estimates to be invalid. Our review of each study did not reveal authors' arguments excluding VSL estimates except a few instances in which authors questioned the reliability of the BLS and NIOSH occupational risk data. Because it is accepted to use these risk data in hedonic wage studies, we did not view this as a valid reason for dropping those VSL estimates. The summary of each author's review of their VSL estimate is in an appendix available upon request from the authors.

<sup>8</sup> Most studies use the hourly wage or weekly wage. In those cases authors multiply by 2000 (some use 2080) for mean hourly wage, and 50 (some use 52) for mean weekly wage to obtain mean annual wage. We follow each study's estimation approach and if that is not available, we use a multiplier of 2000 for hourly wage and 50 for weekly wage.

<sup>9</sup> The coefficient  $d \ln Y / d p_i$  does not depend on the units in which Y is measured. The requirement for a comparison across is that results are converted in the same units, e.g. per thousand per year.

<sup>10</sup> To assure the quality of re-estimation of VSL, we matched our results with estimates done by the original authors when available. Although the VSL estimates from Kneisner and Leeth (1991), Smith and Gilbert (1984) and V.K. Smith (1976) are included in EPA 812 report, the original manuscripts do not provide VSL estimates, and we could not replicate the estimates reported in EPA 812. Therefore we exclude those studies from our analysis.

<sup>11</sup> A full listing of studies and their associated VSL are available from the authors upon request.