Environmental & Resource Economics (2006) 34:385–406
DOI 10.1007/s10640-006-9000-8

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

IKUHO KOCHI[1], BRYAN HUBBELL[2] and RANDALL KRAMER[3,]*
[1]*Department of Economics, Georgia State University, Atlanta, GA, USA;* [2]*U.S. Environmental Protection Agency, Research Triangle Park, NC, USA;* [3]*Nicholas School of the Environment and Earth Sciences, Duke University, Durham, NC 27708-0328, USA;* **Author for correspondence (e-mail: kramer@duke.edu)*

Accepted 13 May 2005

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 2002, containing 197 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 influenced by the choice of valuation method, study location, and union status of sample but not to the source of data on occupational risk or the consideration of non-fatal risk injury.

Key words: value of a statistical life (VSL), empirical Bayes estimate, environmental policy, health policy, contingent valuation method, hedonic wage method

JEL Classification: 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; 2005), 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

★ Without implicating them for remaining errors, the authors thank Joe Aldy, Jim Hammitt, David Higdon, Marty Smith, and Kerry Smith for comments on an earlier draft of this manuscript.

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.[1] 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, which is what society 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 (\$10×100,000). 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 used a mean VSL estimate 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 2002. 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 \$0.2 million (Loomis and du Vair 1993), 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) estimated a VSL of \$10.4 million with standard error of \$5.2 million, while Miller et al. (1997) reported almost the same VSL (\$10.5 million), but with a much smaller standard error (\$1.5 million).[2] As

Marin and Psacharopoulos (1982) suggested, more weight should be given to VSL estimates that have smaller standard errors.[3]

Mrozek and Taylor (2002) and Viscusi and Aldy (2003) conducted meta regression analyses to evaluate the sensitivity of VSL estimates from hedonic wage studies to study design characteristics, including average risk, income level, data source and composition of sample. Based on their regression results, they provided a plausible range of VSL values. However, Viscusi and Aldy did not take into account the precision of individual VSL estimates, while Mrozek and Taylor used sensitivity analysis to reflect the approximate precision of each VSL estimate. 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 empirical Bayes method has been used in meta-analysis of epidemiology and clinical studies (DerSimonian and Laird 1986; Post et al. 2001). However, to our knowledge it has not previously been applied in meta-analyses of economic studies. Also, while Mrozek and Taylor and Viscusi and Aldy focused on the hedonic wage literature, our analysis facilitates the combination and comparison of VSL estimates from both the hedonic wage and contingent valuation literature.

The first stage of our method 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 variance of the individual VSL estimates, as well as the observed variability between VSL estimates from different studies. This differs from previous meta-analyses in assigning a greater weight to VSL estimates measured with greater precision. 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, specification of the wage equation, union vs. non-union samples, the addition of estimates with missing standard errors, use of a 5% trimmed sample and addition of negative VSL estimates. 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-analyses 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.[4] 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 found that all the HP studies evaluated failed to provide an unbiased estimate of the risk-dollar tradeoff, and tended 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 argued 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. While most studies apply a simple regression of the natural log of wage rates on risk levels, a few of the studies estimate the risk-dollar tradeoff using a structural model. However, the complex estimation procedure makes the VSL estimates 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 argued 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.[5]

There are several other selection criteria that are implicit in the 1992 Viscusi analysis.[6] 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) and Viscusi and Aldy (2003) have found that income level has a significant impact on VSL. Because we are seeking a VSL applicable to U.S. policy analysis, inclusion of VSL estimates from low-income countries may bias VSL downward. Income in the U.S. has also changed over time. However, the focus of this analysis is on the estimates as reported in the studies adjusted for inflation. We did not attempt to hypothesize adjustments to reported VSLs for changes in real income. It is not clear what the impact of such adjustments would be. However, other recent meta-analyses, such as Viscusi and Aldy (2003), have found a positive but inelastic relationship between VSL and income. 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.[7] 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 that have 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-estimated 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).[8] 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 later present a sensitivity analysis showing the effects of excluding the negative estimates.

1.2.1. *Estimation of VSL from HW studies*

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

$$\ln Y_i = a_1 p_i + a_2 q_i + a_3 p_i^2 + X_i \beta + \varepsilon_i \tag{1}$$

where Y_i is equal to earnings of worker i, p_i and q_i are occupational fatal and non-fatal risk faced by i, X_i is a vector of other wage determinants including a constant term, and ε_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.

$$\text{VSL} = (\partial \ln Y / \partial\, dp_i) \times \text{mean annual wage}^9 \times \text{unit of fatal risk}^{10} \tag{2}$$

Note that $\partial \ln Y / \partial p_i$ 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"×"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.

1.2.2. *Estimation of standard error of VSL from HW studies*

Calculation of the variances for the set of VSL estimates from the literature is a key element for the empirical Bayes analysis. While we would ideally like to calculate the complete variance for each VSL estimate, we were unable to do so due to missing data on elements of the variance equation. For example, because most authors calculate the VSL using the mean wage for the sample, the variance of the VSL from a HW study is

$$\text{Var(VSL)} = (\text{unit of risk})^2 \text{Var}(\partial \ln Y / \partial p \times Y) \tag{3}$$

It is typically evaluated at $Y = \overline{Y}$, where $\overline{Y}$ is the average wage for the sample. However, the VSL might also be calculated using the mean predicted wage from the wage equation, which would imply a different variance formula. Or, the

VSL for each observation in a sample could be calculated and then mean VSL computed over the sample. Because most authors did not provide the mean predicted wage, we have chosen to calculate all VSL for this analysis using the mean observed wage. The variances, and thus the weights in the empirical Bayes analysis, are then driven primarily by the variance of the estimated risk parameter, rather than by the overall variance of the wage equation.

For example, if the wage equation is specified as

$$\ln Y = a_1 p_i + a_2 q_i + a_3 p_i^2 + a_4 p_i \mathrm{UNION} + \varepsilon_i, \tag{4}$$

then

$$\begin{aligned}\mathrm{Var(VSL)} &= (\mathrm{unit\,of\,risk})^2 \\ &\times \left[\begin{array}{l}\mathrm{Var}(a_1 Y) + 4\mathrm{Var}(a_3 pY) + \mathrm{Var}(a_4 Y \cdot \mathrm{UNION}) + 4\mathrm{Cov}(a_1 Y, a_3 pY) \\ +4\mathrm{Cov}(a_3 pY, a_4 Y \cdot \mathrm{UNION}) + 2\mathrm{Cov}(a_1 Y, a_4 Y \cdot \mathrm{UNION})\end{array}\right]\end{aligned} \tag{5}$$

where Y, p, and UNION are evaluated at their sample means.

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

$$\mathrm{Var}(a_1 Y) = Y^2 \frac{s^2(a_1)}{n} + a_1^2 \frac{s^2 Y}{n} - \frac{s^2(a_1) s^2(Y)}{n^2} \tag{6}$$

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

$$\begin{aligned}\mathrm{Var(VSL)} =& (\mathrm{unit\,of\,risk})^2 \Big[\overline{Y}^2 \mathrm{Var}(a_1) + 4\overline{p}^2 \overline{Y}^2 \mathrm{Var}(a_3) \\ &+ \overline{Y^2}\, \overline{\mathrm{UNION}^2} \mathrm{Var}(a_4)\Big]\end{aligned} \tag{7}$$

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 40 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. This does not however, address the larger issue of whether the use of mean observed wage in the place of mean predicted wage would have a significant impact. Because of the lack of information on mean predicted wage, we are unable to examine the impact of this assumption.

1.2.3. *Estimation of VSL and standard error from CV studies*

For most of the CV studies, 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. The CV studies that do not report standard errors are removed from our analysis. The effects of excluding these CV studies are examined in a sensitive analysis.

1.2.4. *Estimation of representative VSL for each study*

Most studies reported multiple VSL estimates. For the empirical Bayes approach used in our analysis, each estimate is assumed to be an independent sample (Levy et al. 2000). To assure the independence of each estimate given the multiple observations from single studies, 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

$$Q = \sum w_i(\beta_i - \beta^*)^2 \tag{8}$$

where β_i is the reported VSL from study, $\beta^* = \frac{\sum w_i\beta_i}{\sum w_i}$, and $w_i = \frac{1}{s_i^2}$, where s_i^2 is the sample variance.

Under the null hypothesis of homogeneity, Q is approximately a χ^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 and the standard error for the subset 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

The empirical Bayes estimation technique is a method to improve study-specific estimates by adjusting estimates by within and between study variability. While simple variance weighting is a common approach to incorporating precision of individual estimate in the analysis, under the assumption that the all estimates belong to one underlying distribution, the empirical Bayes method provides an estimator with smaller mean squared error than an estimator that is adjusted by only within study variability (Post et al. 2001).

Following DerSimonian and Laird (1986) and Post et al. (2001), the empirical Bayes model is constructed in a following way. The empirical Bayes model assumes that

$$\beta_i = \mu_i + e_i \tag{9}$$

where β_i is a specific VSL estimate i, μ_i is the true VSL, e_i is the sampling error (within study variability) with a variance s_i^2. The model also assumes that

$$\mu_i = \mu + \delta_i \tag{10}$$

where μ is the mean population estimate and δ_i captures the between study variability that is τ^2.

The weighted average of the reported β_i is described as μ_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 β's (τ^2). These are expressed as follows:

$$\mu_w = \frac{\sum w_i^* \beta_i}{\sum w_i^*} \tag{11}$$

$$\text{s.e. } (\mu_w) = \left(\sum w_i^*\right)^{-1/2} \tag{12}$$

where $\mathrm{w}_i^* = \frac{1}{w_i^{-1}+\tau^2}$, and τ^2 is estimated as

$$\tau^2 = \max\left(0, \left(\frac{(Q-(n-1))}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}}\right)\right) \tag{13}$$

where Q is Cochran's Q statistics described before.

The adjusted estimate of the β_i is estimated as

$$\text{Adjusted}\,\beta_i = \frac{\frac{\beta_i}{s_i^2} + \frac{\mu_w}{\tau^2}}{\frac{1}{s_i^2} + \frac{1}{\tau^2}} \tag{14}$$

This adjustment, as illustrated in Figure 1, pulls the reported estimates of β_i towards the pooled estimate. The more within-study variability, the less weight the β_i receives relative to the pooled estimate, and the more it gets adjusted towards the pooled estimate. The adjustment also decreases the variance surrounding the β_i by including information from all β's into the estimate of β_i (Post et al. 2001).

To compare the different distributions of VSL, we applied the bootstrap method, which is a non-parametric estimation approach. Bootstrapping is equivalent to random sampling with replacement and 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.

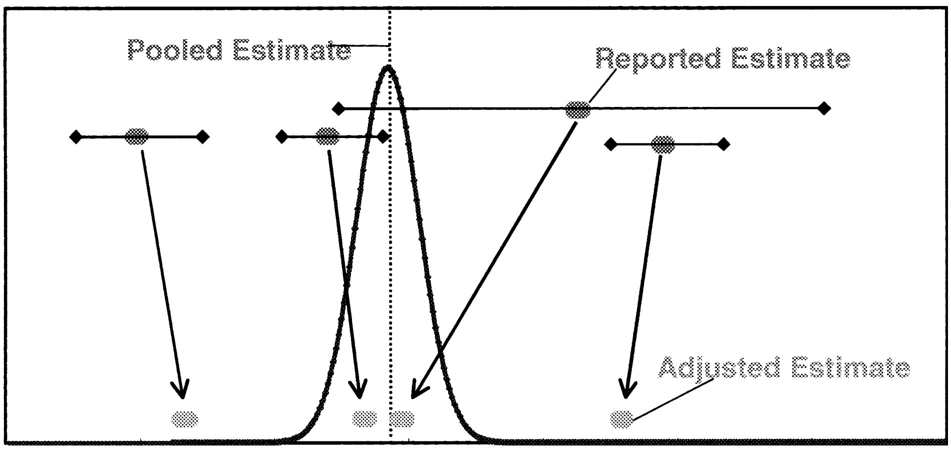

Figure 1. Illustration of empirical Bayes pooling.

2. Results and Comparisons

In total, we collected 47 HW studies and 29 CV studies. A data summary for each stage of analysis is shown in Table I. 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 2002. We re-estimated all possible VSL for the selected studies, and obtained 232 VSL estimates.[11,12] There were five studies that contained negative VSL estimates or VSL estimates for which standard errors were not available, reducing the number of studies used for the main analysis to 40. These studies are excluded from our primary analysis, although we examine the impact of excluding those estimates 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 197 VSL estimates show a mean of \$10.8 million, an extremely wide range from \$0.1 million to \$95.5 million, and a coefficient of variation of 1.3 (in 2000 constant dollars). The VSL estimates from the 60 subsets have a mean of \$10.0 million and a range from \$0.3 million to \$76.0 million with a coefficient of variation of 1.2. The adjusted VSL estimates have a mean of \$5.4 million and a range from \$0.7 million to \$13.9 million with a coefficient of variation of 0.4. Paired t-tests show that the empirical Bayes approach yields a statistically different mean VSL as compared to a simple arithmetic average VSL estimate from the 60 subsets at the 95 percent confidence level.

Table I. VSL data summary

	HW	CV	Total
Number of collected studies	47	29	76
Number of selected studies	31	14	45
Number of estimated VSL	181	51	232
Number of positive VSL with imputed SE	162	35	197
Mean (million \$) (coefficient of variation)	12.3 (1.2)	3.9 (0.6)	10.8 (1.3)
Number of VSL subsets at 1st stage	42	18	60
Mean (million \$) (coefficient of variation)	12.7 (1.1)	3.8 (0.8)	10.0 (1.2)
Number of VSL subsets at 2nd stage	42	18	60
Empirical Bayes mean (million \$) (coefficient of variation)	9.6 (0.5)	2.8 (0.5)	5.4 (0.4)

2.1. THE COMBINED DISTRIBUTION OF VSL

Figure 2 shows 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.[13] The summary results are shown in Table II. The composite distribution of adjusted VSL has a mean of $5.4 million with a standard deviation 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. COMPARISON OF STUDY SUBSETS

2.2.1. *Comparison of valuation method subsets*

Many researchers argue that the VSL is sensitive to underlying study characteristics (Viscusi 1992; Carson et al. 2001; 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 a bootstrapping method to test the hypothesis that HW and CV estimates of VSL are from the same underlying distribution.

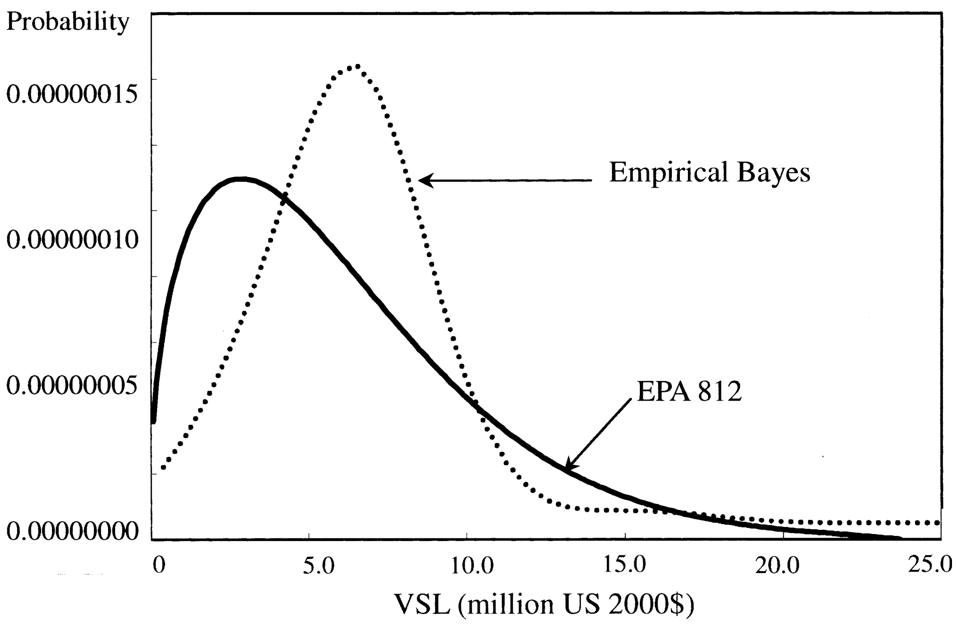

Figure 2. Composite distribution of empirical Bayes adjusted VSL and distribution of VSL based on EPA Section 812 report estimates.

Table II. 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
Distribution comparison by valuation method						
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	0.007
HW (42)	9.6	4.9	0.5			
Distribution comparison by study location (HW only)						
USA (30)	8.9	5.3	0.6	*p*-value (Ho: US = UK)		
UK (7)	22.6	4.9	0.2	< 0.001	< 0.001	0.433
Distribution comparison by occupational risk data source (HW only)						
BLS (20)	10.9	4.9	0.4	*p*-value (Ho: BLS = NIOSH)		
NIOSH (3)	7.2	3.9	0.5	0.151	0.321	0.265
Distribution comparison by non-fatal dummy (NFD) variable (HW only)						
Total (46)	9.1	4.4	0.5	*p*-value (Ho: With NFD = without NFD)		
With NFD (16)	11.0	6.2	0.6	0.212	0.298	0.317
Without NFD (30)	8.3	3.7	0.4			
Distribution comparison by union status (HW and U.S. only)						
Total (10)	13.3	8.5	0.6	*p*-value (Ho: Union = Non-union)		
Union (7)	17.0	9.0	0.5	0.003	0.006	0.040
Nonunion (3)	6.8	0	–			
Distribution comparison by valuation method after adding excluded estimates						
Total (69)	4.7	2.3	0.5	*p*-value (Ho: HW = CV)		
CV (23)	2.6	1.3	0.5	< 0.001	< 0.001	0.009
HW (46)	8.9	4.8	0.5			
5% trimmed estimate						
Total (54)	5.8	2.6	0.4			
Distribution comparison by valuation method after including negative estimates						
Total (67)	4.1	1.7	0.4	*p*-value (Ho: HW = CV)		
CV (18)	2.7	1.3	0.5			
HW (49)	6.4	3.4	0.5	< 0.001	0.004	0.108

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 composite distributions for HW and CV sample are shown in Figure 3. The HW distribution has a mean value of \$9.6 million with a standard deviation of \$4.9 million while the CV distribution has much smaller mean value of \$2.8 million with a standard deviation of \$1.3 million (see Table II). Bootstrap tests of significance show the VSL based on HW is

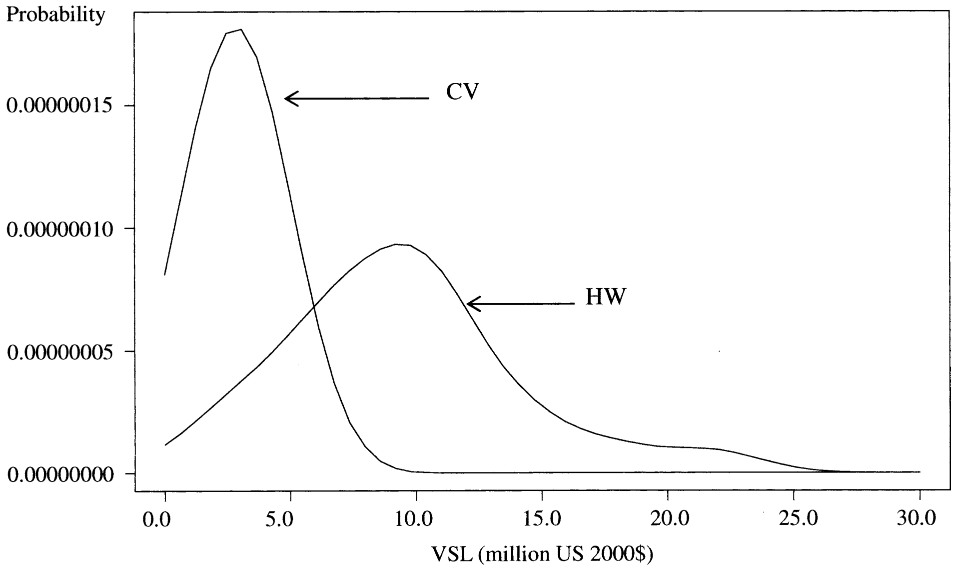

Figure 3. Comparison of composite distributions of empirical Bayes adjusted VSL based on HW and CV estimates.

significantly larger than that of CV ($p < 0.001$), comparing means, medians and interquartile ranges between the distributions.

2.2.2. *Comparison of study location subsets*

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 $8.9 million with a standard deviation of $5.3 million, while the distribution for the U.K. sample has a mean value of $22.6 million with a standard deviation of $4.9 million. Bootstrap tests of significance show that the U.S. estimates are significantly different from UK estimates based on a comparison of means and medians between distributions.[14]

2.2.3. *Comparison of source of occupational risk data*

Moore and Viscusi (1988) found that VSL was influenced by 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. The distribution for the BLS sample has a mean value of $10.9 million with a standard deviation of $4.9 million. The distribution for the NIOSH sample has a mean value of $7.2 million with a standard deviation of $3.9 million. The empirical Bays method reduces the variance of distributions,[15] but 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 NIOSH risk data.

2.2.4. *Comparison of fatal and non-fatal injury risks*

Another test is conducted to examine the effect of including a non-fatal dummy variable in HW studies. Previous reviews of VSL literature have noted the importance of accounting for non-fatal injury risk (Viscusi 1992; Viscusi and Aldy 2003). In comparing VSL distributions, we found no significant difference between the distributions with and without the non-fatal risk variable. Although Viscusi (1978) reported a relatively large drop in VSL when a non-fatal injury risk dummy was included, Leigh and Folson (1984) reported only slight reductions in VSL, and Martinello and Meng (1992) reported slight increases in VSL when controlling for non-fatal risk. Mrozok and Taylor's meta analysis also found an insignificant effect of controlling for non-fatal risk.

2.2.5. *Comparison of union and non-union samples*

There has been considerable interest in the effect of union status on VSL in the HW literature (Dickens 1984; Sandy and Elliott 1996; Siebert and Wei 1994). We compared distributions based on union and non-union samples. Most studies control for the influence of union status by including a dummy variable in their HW regression. Therefore, there are a limited number of studies that estimate HW models with only a union or non-union sample. A comparison of studies from the US and UK did not yield a significant difference.[16] When we limited the sample to U.S. studies, we found that the mean VSL for the union sample is $17.0 million, which is significantly larger than the mean VSL for the non-union sample of $6.8 million.

2.3. SENSITIVITY ANALYSIS

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 we estimated 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 deviation of \$2.3 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 studies, which tend 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 estimate 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 \$9.6 million to \$6.4 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. Our results are generally in line with Viscusi and Aldy's findings based on a meta-analysis of hedonic wage studies only. They found that mean predicted VSL from their full sample (including international studies) was between \$5.0 million and \$6.2 million for various specifications. Our mean VSL for the U.S. sample of \$8.9 million is somewhat higher than the Viscusi and Aldy's most comprehensive robust estimate of \$7.6 million for their U.S. sample. Our estimates are higher than the "best practice" VSL of \$2 million found by Mrozek and Taylor (2002). This difference comes from their adjustment for inter-industry differences in compensating wage equations.

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 carefully consider valuation methods 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).

It is not surprising that such differences have been found. Theoretically, the two valuation methods do not necessarily provide the same results for non-marginal risk changes, because "… the revealed preference approach is estimating a local trade-off, while the contingent valuation approach approximates a movement along a constant expected utility locus" (Lanoie et al. 1995, p. 236). Also, the CV and HW methods use different populations and different types of risks, both of which can lead to a systematic disparity in VSL estimates. It is also important to consider the bias that may stem from each evaluation method. For example, CV studies may suffer from hypothetical bias (Cummings and Taylor 1995), while HW studies may be subject to bias resulting from measurement errors (Black 2001), and omitted variables (Garen 1988; Hwang et al. 1992; Gunderson and Hyatt 2001). However, the impact and direction of the differences across methods have not been systematically investigated using aggregated data prior to this analysis.

Our sensitivity analysis found no significant difference on average in the VSL estimates from studies using BLS vs NIOSH data. Additional research into appropriate measures of risk for VSL studies is needed. 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 2003).

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 systems may be a key influence, 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. (2002) 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 carefully by assessing any potential questionnaire or 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.

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 197 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 197 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. In future meta-analyses, methods will also need to be developed to control for possible dependence among VSL estimates derived from studies using similar databases or from the same authors. We have used a relatively simple grouping procedure based on the Cochran's Q-test for homogeneity for a given author. However, there may be correlation between VSL estimates even across authors when similar datasets are used. This may lead to unwarranted weight being placed on VSL estimates all derived from a single data source.

The widely cited estimate of $6.3 million from the EPA 812 study based on Viscusi's assessment of the VSL literature was a simple average giving equal weight to each estimate. This early approach ignored within and between study variability. Mrozek and Taylor (2002) 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 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.

Notes

1. 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.
2. 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.
3. As a reviewer pointed out, this weighting argument is true if there are no omitted, confounding factors. For example, hedonic wage studies that omit a non-fatal injury risk variable often produce larger and more precise estimates for the fatal injury risk coefficient than studies that include a non-fatal risk variable (Viscusi and Aldy 2003). We examine the impact of including a non-fatal risk variable in a comparison of subsets in Section 2.2.4.
4. We also employed fixed effects approaches for pooling, but found this resulted in an artifact of providing greater weight to studies whose authors reported multiple estimates.
5. 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.
6. 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–$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.
7. From http://worldbank.org/data/databytopic/class.htm. High-income OECD members have annual income greater than $9266 per capita.
8. 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.

9. 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.
10. The coefficient $\partial ln\ Y/\partial dp_i$ 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.
11. 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 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.
12. A full listing of studies and their associated VSL are available from the authors upon request.
13. To visualize the composite distribution, we used a kernel density estimation approach to develop smoother distributions than the histogram approach. See Silverman (1986) for more details about this approach. We arbitrarily chose a bandwidth equal to 0.7, but this choice does not affect any of our analytical results in Sections 2 and 3.
14. Excluding VSLs based on low-risk occupations in UK sample does not change this result.
15. The simple mean of the NIOSH sample is \$7.1 million with a standard deviation of \$5.0 million and the mean of the BLS sample is \$10.9 million with a standard deviation \$7.3 million.
16. This may be because the compensation systems for workplace accidents for unionized and non-union workers are quite different in the U.S. and U.K. (Siebert and Wei 1994).

References

Anderson, G. F. and P. S. Hussey (2000), *Multinational Comparison of Health Systems Data 2050*. The Commonwealth Fund, International Health Policy Report (October 2000).

Arabsheibani, R. G. and A. Marin (2000), 'Stability of Estimates of the Compensation for Damage', *Journal of Risk and Uncertainty* **20**(3), 247–269.

Black, Dan A. (2000), '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.

Carson, R. T., N. E. Flores and N. F. Meade (2001), 'Contingent Valuation: Controversies and Evidence', *Environmental and Resource Economics* **19**(2), 173–210.

Carson, R. T. (1996), 'Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods', *Land Economics* **72**(1), 80–99.

Cummings, R. G. and L. O. Taylor (1995), 'Unbiased Value Estimates for Environmental Goods: a Cheap Talk Design for the Contingent Valuation Method', *American Economic Review* **83**(3), 649–665.

DerSimonian, R. and N. Laird (1986), 'Meta-analysis in Clinical Trials', *Controlled Clinical Trials* **7**(3), 177–188.

Dickens, W. T. (1984), 'Differences between Risk Premiums in Union and Nonunion Wages and the Case for Occupational Safety Regulation', *American Economic Review* **74**(2), 320–323.

Eom, Y. S. (Nov 1994), 'Pesticide Residue Risk and Food Safety Valuation: A Random Utility Approach', *American Journal of Agricultural Economics* **76**, 760–771.

Fisher, A., L. G. Chestnut and D. M. Violette (1989), 'The Value of Reducing Risks of Death: A Note on New Evidence', *Journal of Policy Analysis and Management* **8**(1), 88–100.

Garen, J. (1988), 'Compensating Wage Differentials and the Endogeneity of Job Riskiness', *Review of Economics and Statistics* **70**(1), 9–16.

Goodman, L. A. (1960), 'On the Exact Variance of Products', *American Statistical Association Journal*, December, 708–713.

Gunderson, M. and D. Hyatt (2001), 'Workplace Risks and Wages: Canadian Evidence from Alternative Models', *Canadian Journal of Economics* **34**(2), 377–395.

Hammitt, J. K. and J. D. Graham (1999), 'Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability?', *Journal of Risk and Uncertainty* **8**(1), 33–62.

Hwang, H.-S., W. R. Reed and C. Hubbard (1992), 'Compensating Wage Differentials and Unobserved Productivity', *Journal of Political Economy* **100**(4), 835–858.

Kneisner, T. J. and J. D. Leeth (1991), 'Wage Differentials for Fatal Injury Risk in Australia, Japan, and the United States', *Journal of Risk and Uncertainty* **4**(1), 75–90.

Krupnick, A., A. Alberini, M. Cropper, N. Simon, B. O' Brien, R. Goeree and M. Heintzelman (2002), 'Age, Health, and the Willingness to Pay for Mortality Risk Reductions: A Contingent Valuation Survey of Ontario Residents', *Journal of Risk and Uncertainty* **24**(2), 161–186.

Lanoie, P., C. Pedro and R. Latour (1995), 'The Value of a Statistical Life: A Comparison of Two Approaches', *Journal of Risk and Uncertainty* **10**(3), 235–257.

Leigh, P. J. and R. 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), 55–56.

Levy, J. I., J. K. Hammitt and J. D. Spengler (2000), 'Estimating the Mortality Impacts of Particulate Matter: What Can Be Learned from Between-Study Variability?', *Environmental Health Perspectives* **108**(2), 108–117.

Loomis, J. B. and P. H. du Vair (1993), 'Evaluating the Effect of Alternative Risk Communication Devices on Willingness to Pay: Results from a Dichotomous Choice Contingent Valuation Experience', *Land Economics* **69**(3), 87–98.

Manly, B. F. J. (1997), *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd edition. London: Chapman & Hall.

Marin, A. and G. Psacharopoulos (1982), 'The Reward for Risk in the Labor Market: Evidence from the United Kingdom and Reconciliation with Other Studies', *Journal of Political Economy* **90**(4), 827–853.

Martinello, F. and M. Ronald (1992), 'Workplace Risks and the Value of Hazard Avoidance', *Canadian Journal of Economics* **25**(2), 333–345.

Miller, T. R. (2000), 'Variations between Countries in Values of Statistical Life', *Journal of Transport Economics and Policy* **34**(2), 169–188.

Miller, P., C. Mulvey and K. Norris (1997), 'Compensating Differentials for Risk of Death in Australia', *Economic Record* **73**(223), 363–372.

Moore, M. J. and W. Kip Viscusi (1988), 'Doubling the Estimated Value of Life: Results Using New Occupational Fatality Data', *Journal of Policy Analysis and Management* **7**(3), 476–490.

Mrozek, J. R. and L. O. Taylor (2002), 'What Determines the Value of Life? A Meta-Analysis', *Journal of Policy Analysis and Management* **22**(1), 253–270.

Post, E., D. Hoaglin, L. Deck and K. Larntz (2001), 'An Empirical Bayes Approach to Estimating the Relation of Mortality to Exposure to Particulate Matter', *Risk Analysis* **21**(5), 837–842.

Sandy, R. and R. F. Elliott (May 1996), 'Unions and Risks: Their Impact on the Level of Compensation for Fatal Risk', *Economica* **63**, 291–310.

Scotton, Carol R. and Laura O. Taylor (2000), 'New Evidence from the Labor Markets on the Value of a Statistical Life', Working Paper, Atlanta: Department of Economics, Andrew Young School of Policy Studies, Georgia State University.

Siebert, S. W. and X. Wei (1994), 'Compensating Wage Differentials at Workplace Accidents: Evidence for Union and Nonunion Workers in the UK', *Journal of Risk and Uncertainty* **9**(1), 61–76.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall.

Smith, R. S. (1976), *The Occupational Safety and Health Act: Its Goals and Achievements*. Washington: American Enterprise Institute.

Smith, V. K. and C. C. S. Gilbert (1984), 'The Implicit Risks to Life: A Comparative Analysis', *Economic Letters* **16**(3–4), 393–399.

United States Environmental Protection Agency (USEPA) (1999), *The Benefits and Cost of the Clean Air Act, 1990 to 2010*. Washington: USEPA.

United States Environmental Protection Agency (USEPA) (2005), *Regulatory Impact Analysis for the Final Clean Air Interstate Rule*. Washington: USEPA.

Viscusi, K. W. (1978), 'Labor Market. Valuations of Life and Limb: Empirical Evidence and Policy Implications', *Public Policy* **26**(3), 359–386.

Viscusi, W. K. (1992), *Fatal Tradeoffs: Public and Private Responsibilities for Risk*. New York: Oxford University Press.

Viscusi, W. K. and J. E. Aldy (2003), 'The Value of Statistical Life: A Critical Review of Market Estimates throughout the World', *Journal of Risk and Uncertainty* **27**(1), 5–76.

Viscusi, W. K. and W. N. Evans (1990), 'Utility Functions that Depend on Health Status: Estimates and Economic Implications', *American Economic Review* **80**(3), 353–374.