DE GRUYTER DOI 10.1515/jbca-2012-0015 — Journal of Benefit-Cost Analysis 2013; 4(1): 39–80

Carol R. Scotton*

New risk rates, inter-industry differentials and the magnitude of VSL estimates

Abstract: The Census of Fatal Occupational Injury (CFOI) provides data for creating objective measures of workplace risk used in estimations of wage-risk premia for value of statistical life (VSL) calculations. This relatively new data set enables a more theoretically defensible measure for use in hedonic wage equations. However, constructing these rates from the CFOI data necessarily involves creating an industry-occupation matrix defining the "jobs," deciding whether or not to include the self-employed, and selecting a denominator. These choices in the construction of the risk measure alone, as shown here, result in variations of VSL estimates ranging from $8 million to $18 million. Further, risk measures based on the CFOI data, regardless of construction, are sensitive to simple changes in the specification for the hedonic wage equation. In particular, fixed effects describing the industry in which a worker is employed, as well as the worker's occupation, are primary influences on the magnitude of the VSL estimates.

Keywords: compensating wages; hedonic wage model; inter-industry differentials; value of a statistical life; workplace fatalities.

JEL codes: J17; J31; I1.

***Corresponding author: Carol R. Scotton**, Department of Economics, Knox College, Galesburg, IL 61401, USA, Tel.: (309) 341-7115, e-mail: cscotton@knox.edu

1 Introduction

Thomas Schelling (1968), first to describe what has come to be known as the *value of a statistical life* (VSL), made the case for establishing a monetary value of the benefit of life saving policies. Now the VSL is widely used, for example, from the Department of Homeland Security in valuing the reduction of risk from terrorists' attacks in Robinson, Hammitt, Aldy, Krupnick, and Baxter (2010) to a report by Robert Scharff (2010) considering the benefits of food safety. Since its inception, however, the idea of a VSL, as well as the particular point estimate of the VSL, has attracted considerable controversy. The Environmental Protection Agency (EPA) Science Advisory Board was asked to review practices involving the VSL (Cropper and Morgan, 2007). The very notion of a VSL was lampooned on the Colbert Report

(2008) while the term itself prompted discussions for "Euthanizing the Value of a Statistical Life" by Trudy Cameron (2010). The EPA has since issued *Guidelines for Preparing Economic Analysis* where considerable attention is paid to the VSL (2010). The application of and terminology for the VSL are factors that will continue to resonate with the public and in policy discussions. Yet underneath all this the workplace risk rate, fundamental to the predominant method used to estimate the VSL, faces serious methodological concerns which deserve attention.

Comprehensive reviews by Viscusi and Aldy (2003) and Mrozek and Taylor (2002) covering the past 40 years show that most estimates of the VSL are based on labor market data and the relative risk of job related fatalities. For example, Robinson (2007) reports that 21 of the 26 point estimates of the VSL included in the EPA Cost Benefit Analysis of the Clean Air Act (1999) are based on compensating wage differentials and use a variety of measures for the workplace fatal risk rate. Serious concerns about the endogeneity of workplace risk, assumptions regarding risk aversion, effects of non-fatal injury or work-related illnesses and difficulties in generalizing from a sample to the population are well-known and have been discussed, for example, in Ashenfelter (2006) and Ashenfelter and Greenstone (2004) as well as in the reviews cited above.[1] The February 2010 special issue of the *Journal of Risk and Uncertainty*, edited by W. Kip Viscusi, is devoted to considering the heterogeneity of VSL. In this journal, Hammitt and Robinson (2011) explore the role of income elasticity and the VSL. Studies that aim to address these concerns using labor market data must inevitably begin with a measure of fatal risk in the workplace. Yet the VSL, as shown in this study, is highly sensitive – varying between \$11 million to \$18 million – depending *only* upon how the same data is used to construct the risk rate; it varies by even more – from about \$5 million to over \$40 million – depending *only* upon the mix of industry and occupation controls. How these risk rates are constructed and how the construction of the risk rate interacts with the use of industrial and occupational controls necessary to identify the wage-risk premium is the subject of this paper. By making a series of choices in the construction of the rate explicit, guidance in developing more transparent rates is offered.

1.1 Background

A VSL point estimate based on labor market data is derived from hedonic wage equations in which wages are regressed on a fatal risk rate and a set of worker

1 There is a vast literature on the VSL and the debates and concerns about it, prompting specialty websites such as VSL Research at the Maxwell School of Syracuse University found at <http://sites.maxwell.syr.edu/vsl/vsl.html> last accessed August 7, 2012.

and job characteristics.[2] As originally described by Jones-Lee (1974), Thaler and Rosen (1976), and Smith (1979), the theory relies on compensating wage differentials; consequently, estimating these wage-risk premia requires measures of the relative riskiness of the various jobs. Therefore, the risk rates – specifically how they are constructed – bear close attention. A set of "jobs," however, for which to compute these rates is not a given. Intuitively, an electrician, for example, working for a utility company has a different job than one working for a hospital. But there is no comprehensive list of jobs that adequately distinguishes these differences. Instead there are occupational classifications describing tasks or skill sets required of individuals and industrial codes defining the (major) product or service provided by firms. A particular job is not categorized in any systematic or formal way. Instead, jobs, or sets of job categories, are usually defined on an ad hoc basis. So, for example, Gittleman and Pierce (2011), in their study on inter-industry differentials, define jobs by "using the employer's most narrow occupational classification" (p. 359).

Ideally, the risk rate would be derived directly from workers' subjective assessments of the threat of a sudden workplace fatal injury. Using such measures, however, would present a host of commensurability concerns even if they were possible to collect. Instead, the primary literature employs probabilities extracted from the historical record of workplace deaths. Because of data constraints, all fatality rates used until recently – and the VSL estimates included in the reviews and studies mentioned above – are based either on workers' industry alone, or their occupation alone.[3] Indeed, a major criticism in using workplace fatalities was based on this limitation, for example, see Leigh (1995). The primary data available were deaths within industry groups.[4] This meant that regardless of occupation, all workers within a specific industry were assigned the same probability of death. For example, the risk faced by electricians versus that of

2 The VSL represents the minimum amount that the workers would, collectively, be willing to pay to reduce the risk of death by one person. For example, a VSL of $6 million is computed from an estimated annual wage-risk premium of $600 from risk rates based on 1 in 10,000 workers. Each of 10,000 workers, on average, receives an annual compensating differential of $600 for facing the increased risk that one more person among them will be fatally injured on the job. In other words, the VSL is computed the same way willingness to pay estimates for a public good are: individual wage-risk premia are summed vertically.

3 This is true regardless of whether the terminology used refers to "job," "industrial," "workplace" or "occupational" fatalities.

4 Data was also available by occupation group only, i.e., regardless of industry, from the National Institute of Occupational Safety and Health (NIOSH). Moore and Viscusi (1988) and Black and Kniesner (2003) compare risk rates by industry only, produced by the BLS, and by occupation only available from NIOSH.

truck drivers in the utility industry (or electricians and ambulance drivers in the hospital industry) were not differentiated.[5] In addition to this mismeasurement of the risk rate, rates based on industry alone meant that VSL estimates could not control for known inter-industry wage differentials for fear of canceling out the risk premia.[6] That inter-industry differentials are distinct from compensating wages for risky or undesirable work is borne out by Lane, Salmon, and Spletzer (2007). They report that "the industry wage differentials literature has empirically examined and rejected the hypothesis of compensating differentials as an explanation for the wage differentials". (p. 4). So, because the risk of a fatality on the job depends on both one's occupation *and* industry, failure to account for these inter-industry differences in wages necessarily biases the wage-risk premium.

In 1992, the US Bureau of Labor Statistics (BLS) began the Census of Fatal Occupational Injuries (CFOI) signaling a major advancement in the ability to define workplace risk more accurately as reported by Scotton (2000) and Viscusi (2004).[7] With the availability of the CFOI data, new estimates of the VSL are being made, for example, by Viscusi (2004) and by Scotton and Taylor (2011).[8] However, while based on the same data set (the CFOI) the risk rates being used in these studies are computed differently. Not only are the matrices of occupation/industry pairs (to be discussed later) different but also the computation of the numerator (which deaths are counted) and the denominator (labor force estimates) differ.

5 The common example of coal miners and secretaries obscures the point that both men and women face risks in the workplace and men and women face the same risk if they are doing the same job. Because risk rates were primarily only at the industry level and most fatalities were male, women were routinely excluded from the labor force samples in VSL estimations. Importantly however, female deaths were likely included in the computation of the risk rates. In other words, if the coal miner was a woman, it is still the death of a coalminer. This would have increased the risk rates for the labor force sample and decreased the wage-risk premiums estimated in the wage equations. This study will not distinguish between men or women holding a particular job. The point is to measure the risk inherent in the job, that is, due to the nature of the job, not if the coal miner, secretary, electrician, doctor or driver is a man or a woman.

6 Leigh found that including industry controls rendered the coefficient on risk not statistically significant. The meta-analysis by Mrozek and Taylor 2002 included 25 (out of 142) estimates where industry dummy variables were included in the wage equation. The presence of industry dummies significantly lowers the VSL estimate.

7 See Drudi (1995) for a comprehensive summary of the history of occupational risk data. The CFOI research data file provides micro-level data on all workplace deaths in the US; importantly, it identifies both the industry and occupation of the worker at the time of death, the circumstances of death, employment status and more. Researchers can gain access to the CFOI data file via a confidentiality agreement with the BLS.

8 The BLS continues to produce fatal occupational injury rates, but only by either industry or by occupation and then only for selected industries and occupations. See <http://www.bls.gov/iif/oshnotice10.htm> last accessed August 7, 2012.

The EPA *Guidelines* recommend that studies generating VSL estimates not only provide a clear identification of the source data for the fatal risk measure but also describe how these data are used in computing the occupational risk rate. The objective in this study is not to add another estimate for the VSL but to clarify the decisions which must be made in constructing the risk rate. First, the specific choices, whether made explicitly or not, are identified and alternative sets of the risk rate, based on different choices, are constructed. Then with these alternative constructions of the risk rate the impact on the resulting VSL estimate can be directly examined independent of all other factors.

While the CFOI data makes it possible to construct a more accurate risk rate, its provision of additional details not only facilitates but also demands other significant theoretical considerations and pragmatic issues be addressed. For example, while the Current Population Survey (CPS) estimates about 7% of the labor force to be self-employed the CFOI reports 21% of workplace deaths were among the self-employed. This disproportionate share of workplace deaths suffered by the self-employed questions including them in a risk rate used for labor market estimates which are necessarily based on a sample of wage earners only. In addition, it is now possible to explore how choices made in defining a job, when risk rates are based on industry *and* occupation, interact with the level of aggregation used for industrial and occupational controls. Isolation of these alternative choices made in the construction of the risk measure definitively illustrates that the magnitude of the VSL estimate is sensitive to 1) how the risk rate is constructed, and 2) which industrial and occupational controls are used in the wage equation.

The next section describes the construction of the occupational fatal risk rate with the data now available and presents six possible variations for computing the risk rate (see Figure 1). Then, using a standard labor force sample and empirical model, I illustrate how the risk rate construction by itself impacts the magnitude of the wage-risk premia. Next, I show the effect of varying the level of aggregation of industrial and occupational controls used in the model. A discussion and conclusion complete the paper.

2 Constructing a workplace fatality risk rate

An objective measure for a workplace fatality risk rate is the incidence representing the number of workplace deaths recorded for a specific “job” per 10,000 workers in that “job” and is calculated as:

$$risk_{oi} = \frac{D_{oi}}{W_{oi}} \times 10{,}000 \tag{1}$$

where $risk_{oi}$ is the risk rate for the "job" defined by occupation, o, and industry, i (an occupation/industry pair); D_{oi} is the annual average number of deaths in that occupation/industry pair; and W_{oi} is the annual average number of workers in the same occupation/industry pair.[9] This equation, while seemingly very simple and straightforward, conceals critical choices embedded within it. The specifications for the count of D_{oi} and an estimate of W_{oi} require translating the available data representing a complex workforce in order to construct a discrete set of relative risk rates. Foremost is assuring compatibility of the information on occupation and industry among the data sources for D_{oi} and W_{oi} as well as with labor force data sets that could be used for VSL estimation. Three broad choices must be made.[10] First, is establishing parameters for the matrix of occupation/industry pairs which defines a set of jobs; next, is whether to include all classes of workers or only those earning wages and salary; and finally, is selecting which data source to use for estimating the denominator.[11] Each of these factors represents a systematic non-sampling error in the measure of the relative riskiness of the job. The size of this error, or the direction of the bias in a VSL estimate that uses these measures, is impossible to predict. What can be examined, nonetheless, is how decisions made in the construction of the risk rate impact the VSL estimate. This study constructs six versions of the workplace risk rate to make such a comparison.

9 An annual average over a series of years for deaths and the size of the labor force is generally used to smooth out anomalies given the rarity of deaths and the freak, and unfortunate, occurrence of multiple deaths in a particular occupation/industry pair happening in one event.

10 There are other choices that could be made to construct more specific rates, for example, by gender, by geographic region or by length of experience on the job (if it were possible to obtain valid estimates of W_{oi}). In any case, decisions regarding the three factors considered in this study would have to be made; the finer distinctions would just make the task that much more challenging. Further, and more usefully, the risk rates can be constructed to differentiate the type of risk faced, such as transportation-related vs homicides [see, for example, Scotton and Taylor (2011) or Kochi and Taylor (2011)].

11 Beginning in 2008, the BLS is computing fatal injury rates based on hours worked, rather than employment-based rates (Northwood, 2010). This changes how the denominator is calculated but does not resolve matters concerning the source of the denominator data. The BSL published rates are still computed only either by industry or by occupation, regardless if hours based or employment based. Because of data limitations across my comparisons, I only use employment-based rates in this study.

2.1 Establishing the matrix of occupation and industry "job clusters"

The relative riskiness of each job has to be established in order to assign a fatal risk rate to workers in a labor force sample. But there is no definitive catalog of jobs to refer to, nor does it seem possible to identify each and every manner of working for listing a complete set of discrete jobs. So it falls upon researchers to determine a specific configuration that reasonably represents a comprehensive set of "job clusters." To do this, labor related research relies on two separate sets of categories: one comprised of occupations and one of industries. Much research, including past VSL research, confine analysis to either occupation or industry. Now, however, with the advent of the CFOI, occupational categories can be paired with industrial categories to form a matrix of "job clusters" for identifying their relative risks. The number of job clusters – henceforth, more simply referred to as jobs – determine the number of possible different risk rates. Constructing a matrix of these occupation/industry pairs, however, proves to be a complex exercise involving numerous practical decisions. All the while, it must be kept in mind that two data sets are used to create the risk rate and the resulting rate must be coded in such a way as to conform to a third data set with the labor force observations.[12] An exploration of some of the choices involved follows.

Robust matrices of occupation/industry pairs can be formed using two readily available classification systems: the North American Industrial Classification System (NAICS) and the Standard Occupational Codes (SOC). These classification systems, it should be noted, while not designed specifically to identify "jobs," are widely used in labor force studies and provide a useful proxy for our purpose.[13] It is tempting to create a rate for every possible occupation within each industry. But the hierarchical character of the coding conventions used in these classification systems drill down to details that may lend very little to the task of differentiating job riskiness. While the number of deaths (D_{oi}) in a particular occupation/industry pair is a count from the census, the corresponding estimate of the level of employment (W_{oi}) at such detail would be based on very

12 Further, the accuracy and reliability of survey responders and coders producing the datasets are a factor. For example, Mellow and Sider (1983) and Mathiowetz (1992) show there is less correspondence in identifying the occupation and industry category between an employee and the firm as the coding is more detailed. This favors more aggregated industrial and occupational groups.
13 The CPS, the most common source for W_{oi}, uses census coding for industry and occupation. The CPS coding includes a mapping to the SOC and NAICS. This additional crosswalk on coding presents practical hurdles. When the CPS is used, for example in estimating the denominator for the risk rate or as the source for the labor market sample, some coding anomalies have to be addressed, which are not fully discussed here.

small samples for many jobs (occupation/industry pairs). The resulting estimates of the relative riskiness among jobs using such fine toothed distinctions would be unreliable. Instead, the hierarchical coding system could be used to customize the distribution of jobs, calibrating the level of detail deemed appropriate for each occupation and industry category. In addition, there are well established aggregations of occupational and industrial groups, built into the systems as well as those developed by the CPS, for example, that are widely seen in the literature and in reports on labor market activities. While the rough outlines are known to many, closer scrutiny highlights some of the decisions a researcher faces in using these systems to create a matrix of jobs.

The SOC and NAICS start with highly aggregated major categories (the 2-digit level) which are then progressively disaggregated down to 6-digits of specificity. This presents possibilities of considerable variety whether collapsing or expanding an occupational group within a particular industry, or adjusting a grouping of industries to match up meaningfully across a group of related occupations. This flexibility, while useful in the main, can introduce some data irregularities as well if one is not mindful of limits the data itself places on which clusters can be created. For example, within a particular data set, counts (or estimates) available at the 3-digit detail may not add up to those available at the 2-digit detail. And, industry and occupation coding in the CPS data does not map perfectly in detail at the 3-digit level. Further, not all categories are represented according to expected patterns. The Utilities industry, for example, provides no greater detail at the 3-digit level than the 2-digit level; one must go to 4-digits to access further distinctions. The CPS (using census coding) does not disaggregate Construction industries at all, which matters if CPS data is used for the denominator of the risk rate or for the labor force sample. Choices must be made to minimize the impact of these irregularities.

The optimal construction of a matrix for VSL research is the one that best depicts how risk is distributed throughout the labor force. Jobs with little difference in risk require little distinction. Jobs characterized by greater differences in risk ought to be more delineated. Ideally, one would set up a process to review possible occupation groups for each industry group (or vice versa) based on similarity of job tasks and riskiness of the workplaces.[14] This could be practical by working with the 2- and 3-digit level of detail. For example, an expected difference in risk would suggest putting certain categories of professionals, such as legal and computer specialists, into one group, while placing health care professionals into a separate group. Similarly, on the industry side, one could combine most retailers into a single group while distinguishing gas stations as

14 The resulting matrix would have to account for the same level of detail in occupations across all industries. This does not mean that every occupation would be found in each industry.

a separate group. Continuing in this fashion would create unique parings of occupations within industries optimized for VSL estimations. This, in effect, employs the hierarchical structures *carte blanche* without being restricted to pre-determined aggregated groups. The resulting matrix, annotated and transparent, could serve as a standard used by all VSL researchers; ideally, it would also be updated every few years so that the risk measure would continue to reflect current conditions.

This study, however, does not offer an independent construction of such a matrix. The need for constructing an optimal matrix is an institutional one, I believe, and requires an institutional resolution. Instead, I construct two variations following as closely as possible the configurations found in the current VSL literature[15] to illustrate that the choice of matrix design alone does play a role in the wage-risk estimate. Both matrices employ the NAICS and the SOC and define a robust set of jobs based upon common aggregation schemes as published in their coding books or by the CPS. Importantly, both use the same data for death counts and employment levels (the annual average over the 4 years, 2003–2006).[16] The only difference is in how the job matrices are configured. Appendix 1 shows the composition of each matrix, illustrating the different aggregations of occupations and industries used to create the job clusters.

One configuration results in an approximately square matrix consisting of 22 occupations across 20 industries such that all possible pairs are categorized into one of 419 distinct cells, each representing one job.[17] (This is similar to Scotton and Taylor's analysis.) The square matrix consists primarily of the 2-digit major groupings for both occupation and industry. The other one delineates 10 occupations across 76 industries (roughly the 3-digit NAICS) producing a long matrix of

15 All studies to date using the CFOI data are constructed from death data prior to the year 2000. (This includes both Viscusi and Scotton and Taylor referenced in this study.) Over a period of years beginning in 21st century, industry coding in the US has changed from the Standard Industrial Code (SIC) to the NAICS. This coding change affects most all US labor force data. This study uses the new NAICS coding system and therefore uses the current data on deaths (from 2003–2006) and labor force composition. While new coding does not map exactly to the old, the data – and coding – used here will be comparable for studies going forward.

16 Every effort was made to minimize any variation from the standard aggregation categories for each matrix while at the same time not losing available data. However, there are a few deaths in the CFOI (over this period) where industry coding was at the 2-digit level only. Consequently, the long matrix, with the industries at the 3-digit level, creates probabilities based on a total of 22,356, instead of the 22,421 deaths (over 4 years) in the square matrix.

17 Since not all occupations are found in every industry, the number of distinct "job clusters" in a matrix is less than the matrices' dimensions would suggest. In other words, what drives the number of cells in a matrix is the denominator. There can be a job with zero risk (i.e., no deaths), but there is not a job if no one is employed in it.

714 distinct jobs. (This is similar to Viscusi's analysis.) These matrices define jobs differently and therefore assign risk differently. In the square matrix, for example, professional occupations are disaggregated while manufacturing industries are clustered into one. In the long matrix, on the other hand, professional occupations are more aggregated while manufacturing industries are more delineated. In some cases, such as the Utilities and Construction industries, there are no differences in the matrices; and, in both, seven of the occupational categories are the same.

Both approaches presuppose using the coding systems "intact" (to the extent possible), that is, keeping to the established aggregation schemes. While neither is an ideal matrix, these two configurations can illustrate that the matrix choices alone do influence the VSL estimate.

2.2 Determining the class of workers to include in the risk rate

The CFOI research file is an annual census of work-related fatalities among the US labor force.[18] Since the CFOI data includes all deaths resulting from an injury "at work," regardless of worker status, the second decision to make when constructing a fatality risk measure is whether or not to count the deaths of self-employed workers in D_{oi}.[19] The underlying theory posits compensating differentials to be the outcome of implicit negotiations between a worker and a firm, whereby firms can offer either more safety or higher wages. While the self-employed are likely making trade-offs regarding their earnings and the amount of risk they are willing to face, their earnings should not be considered wages derived from a negotiation. Clearly, the wages paid to the electrician employed by a construction company are determined by a very different process than those paid to the

18 The CFOI has established work relationship criteria, including location on or off premises and work-related travel, and all cases are corroborated by two or more independent sources. Someone is "at work" while producing a (legal) product or result "in exchange for money, goods, services, profit or benefit" (US BLS, 2007a). For example, commuting to a job is not in scope; traveling required to do a job is in scope.

19 With the CFOI, deaths among wage earners can be distinguished from those among the self-employed. Personick and Windau (1995) find "several important differences in fatality patterns between the two" (pp. 25–26) Similarly, the CFOI identifies those deaths that are specifically work-related from suicides or deaths stemming from domestic violence or terrorism which also occur while someone is at work. Further, the CFOI research file provides variables on the mode and circumstances of the fatal injury. See Kochi and Taylor (2011) and Scotton and Taylor for a discussion of disaggregated risk rates, that is, rates based on the mode of death. These explorations of heterogeneity in risk are markedly different than considering differences in demographics.

self-employed electrician hired to install a 240W outlet for the clothes dryer in a residence. But both would be in the same occupation/industry pair. For this reason, self-employed workers are excluded from the labor market samples used in estimating compensating wages such as the wage-risk premium. Yet this study is the first time the effect of including the self-employed in the risk rate computation is being explored.

When deaths among the self-employed (and family workers or volunteers) are counted in the computation of the risk rate and there is a systematic difference in the risk faced for a specific job when someone is self-employed as compared to someone employed by a firm, then the computed relative riskiness for that job biases the estimate of the wage-risk premium. A risk rate computed based on all workers in the labor force may be lower or higher than the rate computed based on wage earners alone. There are likely non-random factors depending upon the type of job and whether or not a job might be found outside of the context of a firm. In some cases, the same job (represented by the same occupation/industry pair) may be done by a wage earner or the self-employed, such as electricians. But the risks a self-employed electrician may be exposed to could differ from those of a wage earner, for example, working in isolation, the nature or size of the worksite, etc. That is, the risk depends upon the worker status, not just the job category (the same occupation/industry pair). On the other hand, some jobs, such as steelworkers, are likely only held by employees of a firm. Risk measures ignoring these differences would distort the relative riskiness of various jobs in a labor market sample of wage earners. In the past, this was not a concern since the counts of workplace deaths were reported by firms (but there were many other problems with the counts). The more comprehensive census of deaths provided by the CFOI, however, now requires a conscious decision to either include or exclude non-wage earners in computing the risk rate. To examine the effect of this decision, I produce different risk rates for each of the two matrices of job clusters. One set includes all workers in the computation of the rate (similar to Viscusi) and the other counts only workers earning wages and salary (as in Scotton and Taylor).

2.3 Selecting the data source for the denominator of the risk rate

Since the risk rate is the relative probability of death on the job, an estimate of the total number of workers in a particular job (W_{oi}) is needed.[20] The third choice,

20 This study does not distinguish full-time from part-time work, neither in the death counts or employment levels.

then, is to select which data source can provide the most appropriate denominator for Equation 1.[21] I consider two plausible sources for estimates of the size of the labor force.[22] The CPS, a household survey, provides estimates based on the reported primary jobs of all workers and also indicates each worker's status, whether self-employed or working for wages. If a worker has more than one job, using the CPS for the denominator will undercount the number of workers in those secondary jobs.

The BLS also surveys firms for their employment numbers at each occupation through its Occupational Employment Survey (OES).[23] In this survey, workers employed by multiple firms would be counted in each job they hold. Also, by design, the OES provides estimates only of workers employed at firms. Consequently the decision of which classes of workers to include in the fatality count (described above) also impacts the choice here concerning the source for the denominator. If all workers are to be included in the risk rate computation, the CPS would be the source for the estimate.

While both the CPS and the OES provide estimates of the wage earning labor force, these estimates will differ since the CPS counts only the primary job of each worker and the OES is counting all wage and salary workers in each job. Either estimate of the total number of wage-earners is subject to a great deal of error, both independent of and related to the other two factors described.[24]

In order to examine the consequences of including or excluding the self employed in the fatality counts and choosing one data source compared to the other, I first produce a set of risk rates for a job matrix including all workers using the CPS as the source data for the denominator. Then, restricted to wage and

21 Ruser (1998) proposes the denominator ought to be based on hours worked rather than a count of workers. This is now how the BLS computes rates (Northwood, 2010); the estimates for the hours worked are from the CPS. However, since the necessary data for comparisons examined in this study are not available, I continue to use employment-based estimates. Another denominator possibility proposed by Gittleman and Pierce (2006) is to compute the relative risk in terms of output, that is, real GDP. Their direction for a relative risk measure for a VSL estimate would not be appropriate.

22 Another important quality of the source for the denominator is the coding compatibility for occupations and industries with the CFOI data and the subsequent labor force data set used to estimate the VSL. This is addressed in Section 2.1.

23 See the technical notes in USDL 07-0712 (US BLS, 2007b), for example, for a description of the OES survey (US BLS, 2007c).

24 As reported in Appendix 2, there is a pronounced difference in the distribution of risk based on the choice of denominator, with greater variability from the OES estimates.

salary workers, I create two additional sets of risk rates for each matrix: one with the CPS estimate for the denominator and the other with the OES estimate.

In sum, six measures (risk*) for the fatal occupational risk rate are computed to accommodate the three distinct decisions described above.[25] For ease of exposition, the six measures are named to denote the job matrix, the class of worker included in the computation and the denominator source used. The names indicated in the last column of Figure 1 for each distinct risk* are used in the rest of this paper when referring to a particular measure.[26] The top half of Figure 1 identifies the three measures defined by 419 distinct job clusters within the *square matrix* of occupation and industry pairs; in the bottom half are those defined by

Job clusters		Class of workers included		Source for the denominator		Name for risk measure (risk*)[b]
22 occupations across 20 industries n=419[a]	→	All workers, including the self-employed and family workers	→	CPS	→	Square_All_CPS
	→	Workers earning wages and salary only	→	CPS	→	Square_Wage_CPS
			→	OES	→	Square_Wage_OES
10 occupations across 76 industries n=714[a]	→	All workers, including the self-employed and family workers	→	CPS	→	Long_All_CPS
	→	Workers earning wages and salary only	→	CPS	→	Long_Wage_CPS
			→	OES	→	Long_Wage_OES

[a]See Appendix 1 for a listing of industry and occupation groups.
[b]This is Eq. (1) $risk_{oi} = \frac{D_{oi}}{W_{oi}} \times 10{,}000.$

Figure 1 Variables in construction of a risk measure.

25 All of the measures created are based on the annual average number of deaths – and level of employment—for the US labor force of 2003–2006. The six different sets of fatal risk measures can be made available upon request to the author and through an agreement with the BLS CFOI office after meeting confidentiality requirements.

26 The *oi* designation show in Equation 1 has been dropped for ease of reference since all risk rates used in this study are computed for occupation within industry.

714 distinct job clusters of the *long matrix*. As described earlier, a *job* refers to a particular occupation/industry pair representing a cluster of occupations within an industry group. *Square* and *long* refer to one or the other of the matrices. A comparison of the distribution of workplace risk for the set of jobs for each of the risk* constructions is provided in Appendix 2. Now, given these different measures, each of which provides the relative risk of a complete set of jobs, a direct examination of what impact the choices made in constructing the risk rate have on the resulting wage-risk premium can be done.

3 The labor force sample used to estimate wage-risk premia

The labor force sample is the National Bureau of Economic Research's Merged Outgoing Rotation Group (MORG) from the CPS for the year 2006. The summary statistics for the sample are described in Table 1 and are fairly representative of the full-time, nonfarm, payroll labor market for 16–72 year olds, with the mean age of 40, 46% female, just over 30% with a college degree and mean gross weekly earnings of about $800 ($41,600 per annum, YR2006$). The distribution of the sample by major industry mirrors the current situation where the majority of the sample is employed in Health and Social Services, FIRE[27] and the other service-type industries. About 37% of the workers are in management, professional and related occupations; about 25% in sales and office positions. The larger sample, of 121,608 workers as described in the right-hand columns in Table 1, is used for comparison purposes described in the next section.[28] To examine the effects of the industry and occupational controls, I restrict the sample to workers who report their wages to the CPS, resulting in a sample size of 84,336, as described in the left-hand columns in Table 1.[29] This restriction does not materially

27 FIRE is Finance, Insurance and Real-Estate.

28 No agricultural workers, as designated by the MORG, are in either sample. Also, both samples are restricted to workers in job clusters with annual average employment estimates of at least 1000 workers (as computed for the denominator for the risk rates). This restriction excludes very few workers.

29 See Bollinger and Hirsch (2005) and Hirsch and Schumacher (2004) for a discussion on the use of earning imputation in the CPS. As noted in Table 1, the wage data for 31% of the larger sample has been imputed by the CPS. Bollinger (2001) shows that including workers with imputed wages biases estimates of compensating wage differentials. As a result, Hirsch (2008) suggests "the simplest approach, and not a bad one at that, is to omit imputed earners from the estimation sample" (p. 4). I follow this advice when I have finished making the comparison. This should make my results more relevant for future studies.

Table 1 MORG 2006 LF sample.

	Restricted sample		Unrestricted sample	
Proportion of sample with wages imputed by CPS	0	–	0.31	–
Variables used in regressions	Mean μ (n=84,336)	Std. Dev. σ	Mean μ (n=121,608)	Std. Dev. σ
Dependent:				
Ln weekly wage	6.53	0.55	6.53	0.54
Ln hourly wage	2.79	0.52	2.80	0.51
Hourly wage (earnwke / usual hours)	18.71	10.38	18.75	10.36
Earnwke (gross weekly earnings)	798.50	466.76	796.62	463.64
Risk *, rate per 10,000 workers:[a]				
Square_All_CPS	0.3541	0.6021	0.3596	0.6043
Square_Wage_CPS	0.3365	0.6010	0.3419	0.6036
Square_Wage_OES	0.3995	0.7163	0.4062	0.7226
Long_All_CPS	0.3556	0.6424	0.3604	0.6463
Long_Wage_CPS	0.3388	0.6439	0.3437	0.6486
Long_Wage_OES	0.4022	0.7749	0.4086	0.7803
Covariates:				
Age (regressions also include age^2)	40.50	11.99	40.75	12.12
Female (=1 if individual is female)	0.46		0.46	
Married (=1 if individual is married)	0.58		0.57	
White (=1 if individual is white)[b]	0.85		0.83	
Black (=1 if individual is black)	0.08		0.09	
Othrace (=1 if individual is not black or white)	0.07		0.07	
Hourlypaid (=1 if individual is paid by the hour)	0.56		0.57	
Nohsgrad (=1 if individual did not complete high school)	0.09		0.09	
Hsgrad (=1 if individual graduated from high school)	0.30		0.31	
Somecollege (=1 if individual attended college)[b]	0.29		0.29	
Collgrad (=1 if individual has at least a 4-year college degree)	0.32		0.31	

(**Table 1** Continued)

	Restricted sample	Unrestricted sample
NE (=1 if individual lives in the northeast region)	0.19	0.20
South (=1 if individual lives in the southern region)	0.30	0.31
West (=1 if individual lives in the western region)	0.25	0.24
Midwest (=1 if individual lives in the midwestern region)[b]	0.25	0.24
Under union contract (=1 if individual is covered by a union contract)	0.14	0.13
Bigcity (=1 if individual lives in an MSA with population > 1,000,000)	0.56	0.58
Distribution by industries and occupations	**Percent**	**Percent**
Major Industry		
Mining, Utilities and Construction	10%	10%
Food Processing, Textiles and Equip Manufacturing	15%	15%
Retail and Transportation	19%	19%
FIRE, Telecoms and Publishing	21%	21%
Ed, Health and Social Services	24%	24%
Entertainment and Hotels	7%	7%
Personal, Prof, Civic and Religious Services	4%	4%
Major Occupation		
Management, Professional and Related	37%	36%
Services	12%	12%
Sales and Office	24%	25%
Natural Resources, Construction and Maintenance	12%	12%
Production, Transportation, and Material Moving	15%	15%

[a]All fatal injury rates were generated by the author with restricted access to the BLS CFOI microdata.
[b]Excluded variable in regression.

alter the make-up of the sample as evident in comparing it to the unrestricted sample. The smaller sample represents 330 of the 419 possible jobs defined by the square matrix and 597 of the 714 jobs defined by the long matrix; the larger sample includes 337 jobs in the square matrix and 601 in the long. The mean risk* is approximately 0.4×10^{-4}, or about 1 annual fatality per 25,000 workers, for both samples across all six sets of risk*.

Appendix 2 provides more description on the distribution of each risk*. For example, Panel B shows that more than two-thirds of labor force sample (between 68% and 75% depending upon how risk* is constructed) have risk rates that are estimated to be greater than zero and $<0.25\times10^{-4}$. This indicates that most everyone is in a job that has had at least one work related death during the 4-year period. As expected, since workplace deaths are rare but some jobs are very risky, the risk* are right-skewed in every construction; however, the extremes are different depending upon the matrix and the denominator, and therefore, also differ among all workers and the OES-based estimates for wage-earners only. Panels C and D in Appendix 2 provide details on each risk*, un-weighted by a particular sample.

4 Empirical model

The wage-risk premium used to produce a VSL estimate is estimated by an OLS regression on the hedonic wage equation:[30]

$$\text{wage}_k = \alpha + \beta\, risk^*_k + \sum \lambda\, Z_k + \sum \gamma\, Y_j + \varepsilon \qquad (2)$$

where $risk^*_k$ is a measure of the fatal risk rate given for the kth person's industry and occupation (job).[31] Z_k represents the vector of worker characteristics and Y_j, represents job characteristics. The wage-risk premium is the coefficient on *risk**, β.

4.1 Effect of construction of risk*, alone, on the VSL estimate

Using the different constructions for risk*, as described above, can illustrate how the construction of the risk rate measure impacts the estimate of the VSL. To show

30 The model and theory used here are well-documented. See reviews indentified in the introduction.

31 None of the models described in this study include a non-fatal workplace injury measure because, currently, data on non-fatal injury and illness is only available either by industry or by occupation. Including one or both of these non-fatal rates along with the various industrial and occupational controls adds little and changes nothing about the results of this study. See note 46.

how each particular risk* construction compares to results found in the current literature, Viscusi's 2004 study is replicated. The Viscusi study reports on VSL estimates using a fatal workplace risk by occupation and industry, constructed with the CFOI data. His choices in the construction of the risk measure, as described in his study, are analogous to the measure referred to as Long_All_CPS in Figure 1.[32] With similar model and sample specifications (but using the 2006 labor force sample as described in Table 1), a wage-risk premium is estimated for each risk* as described in Figure 1. The results are reported in Table 2. The dependent variable in the top panel is the natural log of hourly wages and in the lower panel it is the level of hourly wages. The coefficient estimates for β from the Viscusi study are recorded in the first column.

The remaining columns in Table 2 present the coefficient estimates for each risk*. The other explanatory variables, not reported here, have the expected sign and magnitude.[33] The R^2 s are given on the Table and show that the construction of risk*, and in particular whether using the square or the long matrix of jobs, has no effect on the explanatory power of the model. In every case, the coefficient on fatal risk is the expected sign and is statistically significant. (Standard errors for both studies account for heteroscedasticity because of clustering of the jobs into an occupation/industry matrix.[34]) However, the magnitudes of the coefficients *do* differ depending upon which construction of risk* is used. As a reference point, I use F-tests for each risk* against the estimate reported in Viscusi's study (reported in the bottom row of each panel in Table 2). For either dependent variable, the coefficients generated by the measures in the long matrix are a closer fit to the estimate he obtained. The F-tests reveal that for the log-linear model (top panel), Long_All_CPS is the closest match, ($p \geq 0.8313$). For the model where the dependent variable is the level of wages (bottom panel), both Long_All_CPS and Long_Wage_OES are not statistically different from the estimate obtained by Viscusi ($p \geq 0.93$). Given that the risk of death on the job has decreased from the mid 1990's, the period covered by Viscusi's measures, and 2003–2006, the period which the measures in this study represent, as well as expected differences in the labor market from 1997 to 2006, differences in the magnitude of the coefficient

32 Note that Viscusi presents risks in terms per 100,000 workers, while this paper reports values per 10,000 workers. All references to the coefficients from Viscusi in this study are presented in terms per 10,000 workers. The decimal point for the coefficients, and their standard errors, has been moved one place. Also, note that the risks represented in Viscusi's 2004 study are based on workplace deaths during the period of 1992–1997. His MORG labor force sample is from 1997. See the notes at Table 2 for more information.
33 The full regressions results are available at <http://faculty.knox.edu/cscotton/expcover.pdf>.
34 The number of clusters is the number of different jobs (i.e., the matrix for the risk rates) represented in the labor force sample. See notes at Tables 2 and 3.

Table 2 Coefficient estimates for different constructions of risk* using the identical model and labor force sample[a].

	Estimate reported by Viscusi	Square _ All_CPS	Square_ Wage_CPS	Square_ Wage_OES	Long_ All_CPS	Long_ Wage_CPS	Long_ Wage_OES
Dependent variable = natural log(hour wage); n=121608							
$\hat{\beta}^{risk*}$	0.032[b]	0.0431***	0.0491***	0.0364***	0.0307***	0.0348***	0.0302***
(se)	(0.009)	(0.0104)	(0.0100)	(0.0121)	(0.0082)	(0.0086)	(0.0079)
R^2		0.414	0.414	0.414	0.413	0.414	0.414
Implied VSL [d] ($)	11,840,000	15,947,000	18,167,000	13,468,000	11,359,000	12,876,000	11,174,000
Test, risk* =0.032, Prob >F =		0.3007	0.0959	0.7415	0.8313	0.7839	0.7818
Dependent variable = hour wage; n=121,608							
$\hat{\beta}^{risk*}$	0.392[c]	0.6236***	0.7382***	0.5283**	0.4071**	0.4821***	0.4047**
(se)	(0.112)	(0.2096)	(0.1957)	(0.2378)	(0.1686)	(0.1775)	(0.1640)
R^2		0.367	0.367	0.367	0.366	0.367	0.367
Implied VSL[e] ($)	7,840,000	12,472,000	14,764,000	10,566,000	8,142,000	9,642,000	8,094,000
Test, risk*=0.392, Prob >F =		0.2700	0.0778	0.5669	0.9288	0.6120	0.9381

Robust standard errors based on clusters of jobs in parentheses, for the 337 different jobs represented in the sample for the square matrix and the 601 different jobs represented in the sample for the long matrix.[35]
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

[a]The model and sample used in this study follow as closely as possible as reported in Viscusi. Both are a sample of full-time workers from the CPS MORG: for this study, from 2006; the LF sample in Viscusi is from 1997. See Table 1, right hand column, for descriptive statistics of the sample used in this study. Both samples include only civilian nonagricultural workers; exclude workers with wages below $4.75 or those with more than 99 usual hours; and include workers with both reported and imputed wages. The same construction is used for the dependent variables. Both models include nine occupational controls and no industrial controls. However, the sample characteristics and model are somewhat different in a few respects which have no appreciable impact on the comparison.[36]

[b]Viscusi's Table 3 panel A, p. 39. Death risk coef. in model without non-fatal injury and illness rate or expected workers' compensation replacement rate, adjusted for a rate per 10,000.

[c]Viscusi's Table 3 panel B, p. 40. Death risk coef. in model without non-fatal injury and illness rate or expected workers' compensation replacement rate, adjusted for a rate per 10,000.

[d]Implied VSL computed as $\hat{\beta}^{risk*} \times$ mean hour wage $\times 2000 \times 10{,}000$.

[e]Implied VSL computed as $\hat{\beta}^{risk*} \times 2000 \times 10{,}000$.

35 The standard errors reported here from Viscusi are also robust and clustered, but he does not report on the number of jobs (clusters) represented by his sample. His underlying job matrix has 707 cells, compared to a possible 714 in the long matrix in this study.

36 Sample restrictions on age and level of education are slightly different. The Viscusi sample excludes the highest earners. The independent variables used here differ in that they include a control for hourly wage, more levels of education, fewer racial groups and regional controls and no control for public versus private industry. Also, the sample used in this study excludes any jobs, and therefore workers in those jobs, when the number of workers employed in that job are fewer than 1000 nation-wide for any of the matrices.

might have been expected.[37] Yet, while the measures based on a construction similar to his produce nearly similar results to his, the alternate constructions, all else equal, produce *different* results. See, for example, the results for Square_Wage_CPS, in the top panel, where prob >F=0.0959 (and in the bottom panel, prob >F=0.0778).

The practical significance here is that different constructions of the death risk variable produce wage-risk premia ranging from 41 cents an hour to 74 cents an hour (3% to 5% of hourly wages). Converting these to a VSL [computed as the wage-risk premium times 10,000 (workers) times 2000 (hours)] yields a range from approximately $8 million to nearly $15 million in bottom panel (or from $11 to $18 million in top panel). These differences are only attributable to the choices that had to be made in constructing the risk rates since modeling choices and sample characteristics are the same.

Note that the model reported on in Table 2 only controls for the worker's occupation.[38] Turning to the long-standing concern about inter-industry differentials, I next examine how the various measures fare under specifications controlling for industry *and* occupation at various levels of aggregation.

4.2 Models including controls for the fixed effects of industry and occupation

It is commonly expected that workers in different occupations would be compensated differently. Regardless of particular risks one might face on the job, wages are principally determined by the tasks and responsibilities of the job, varying according to skills, education required and length of experience. In addition to these differences based on occupation, there is evidence that wages also vary systematically by industry.[39] What are known as inter-industry differentials describe the observation that establishments, grouped together as a type of industry, tend to pay all workers high (or low) wages across the board.[40] So, for example,

37 Costa and Kahn (2002) argue that VSL estimates, in real terms, would not be constant over time. Their findings suggest that workplace risk rates need to be computed on current injury data for the workforce, rather than considering that an inflation adjustment is sufficient.
38 The Viscusi study replicated here only reported on models with occupation controls. The replications presented in Table 2 also only use nine occupational controls.
39 There are also intra-industry differentials as described in Fairris and Jonasson (2008).
40 While this literature studies establishments, the phenomena is, nevertheless, characterized by industry groups (such as were described by the SIC and now by the NAICS), hence the term, inter-industry differentials. This interchange of the terms *industry* and *establishment* has deep roots, e.g., industrial goodwill and industrial relations.

accountants and janitors working in industry A, while earning wages differently because of their occupation, both earn more than their counterparts in industry B (Lane et al., 2007). These inter-industry differentials can be substantial, and are also present when other factors, including risk and working conditions, are taken into account (Dickens and Katz, 1987; Krueger and Summers, 1987; Thaler, 1989; Goux and Maurin, 1999; Osburn, 2000). These wage patterns reflect two sorts of things. One would be the work environment of the industry, which depends upon the nature of the product or the tools, machinery or materials used. The other would be related to the value of the product or the perceived values (or culture) of the industry. These industry characteristics play a role in wage determination for all occupations. And, like occupational characteristics, they are different from the risk of a fatal injury on the job.

The portions of wages determined by these occupation and industry effects have to be isolated to estimate an unbiased wage-risk premium.[41] The SOC, categorizing occupations by similar skill sets and job responsibilities, and the NAICS, classifying business establishments based on their activity generating their greatest revenue stream, are commonly used to control for these fixed effects in wage equations. However, as with the other factors that ought to be controlled for in a wage equation, there is no established standard or definitive guidance regarding the number of, or degree of aggregation for, these controls. Because this is an open question and the EPA Science Advisory Board (Cropper and Morgan, 2007) recommends that information on covariates such as industry and occupation controls be explained, this study next examines the impact on wage-risk premia by varying the number (and presence) of these controls.

These fixed effects can be controlled for in one of three ways. One way would be to use the exact (un-aggregated) industry and occupation codes as given in the labor force data set. However, the classifications may be unique to a particular labor force sample, making it hard to compare across studies.[42] Another approach might be to mirror the matrix structure used for the risk rates, but the categories of the occupation/industry pairs would not be appropriate for two reasons. First, using the "pairs" would not capture the independent differences for occupation separate from industry. The second, more important, reason is technical: there would be multicollinearity between the controls and the risk rates. However using

41 Mrozek and Taylor's meta-analysis found that the presence of five or more industry controls significantly affects the VSL estimate.

42 For example, the CPS uses census code classifications which are analogous to 3, 4, 5, and, in some cases, 6-digit NAICS or SOC codes. In the restricted sample described in Table 1, there are 240 industry categories and 480 occupation categories; in the unrestricted, 240 industries and 482 occupations.

the particular number and aggregation of occupations and industries separately is one possibility for the third approach. Basing the controls on established hierarchies gives two advantages. The constraints on aggregation discussed earlier[43] when constructing the job matrix for the risk measures do not apply to creating these controls; it is a simple matter of creating indicators in the labor force data set. Further, using defined pre-established hierarchies would allow for consistency and comparison across studies. Industry can be aggregated at five levels of detail as defined in the CPS data dictionary and the NAICS code book, and occupation at three levels of detail as defined by the CPS and SOC documentation.[44]

To compare the effects of the possible choices, dummy variables were created for all these variations for controlling industry and occupation fixed effects. OLS regressions of Equation 2 were then estimated for each of the six constructions of the risk measure, risk* for each variation.[45] In addition to the un-aggregated model including all industry and occupation classifications, this included models representing six levels of aggregation for industry controls (including none, seven, 11, 19, 46, and 73) and four levels for occupations (including none, five, nine, and 21). Models using each of the two matrix categories as controls were also estimated.[46]

For each model, an Akaike information criterion (AIC) and Bayesian information criterion (BIC) was computed (by STATA, version 12.0) post estimation. The AIC and BIC are used to provide an objective goodness-of-fit measure to compare

43 Assuring compatibility between the data used for D_{oi} and W_{oi} requires careful fine tuning so that no data is lost.

44 The NAICS sub-sectors (3-digit code) define the least aggregated industry groups. These 73 industry controls are primarily the same industry groups represented in the long matrix. Slightly more aggregated are the 52 CPS detailed industry recodes, followed by the 19 CPS intermediate groups (similar to industry groups in the square matrix). The CPS major industry groups (closely mapped to the NAICS sectors) are more highly aggregated, with 11 groups. The highest degree of aggregation is the seven 1-digit NAICS groups. For occupations, the most detail is the 21 occupation groups (the same occupation groups as in the square matrix) as defined by the 2-digit SOC major groups. The nine SOC intermediate aggregation groups (the same occupation groups used in the long matrix) are less detailed; these are also the CPS major occupational groups. Finally, the five SOC high-level aggregation groups provide the highest level of aggregation for the occupational controls.

45 Models were estimated that also included non-fatal injury rates, which can only be computed by industry alone or by occupation alone (and have several other flaws), to assure nothing was lost or gained regarding control for industry or occupation. Further results from these models are not presented or discussed here since they yield no additional insight into the main issue on the construction of the fatal risk measure or the effect of industry-occupation controls.

46 To avoid multicollinearity, these models were estimated as follows: When using risk* from the square matrix, the long matrix categories are used as controls; when risk* is from the long matrix, the square matrix categories are used as controls.

these models which vary only by the industry and occupation controls used.[47] In addition, every model was estimated without a fatal risk measure to obtain coefficient estimates on the other regressors, particularly, the industry and occupational controls.

The other independent variables and sample specifications are as described previously and summarized in Table 1. Note, however, that the models, and sample, used to examine these fixed effects are different in two substantial ways from the Table 2 results discussed above. Now, the sample is restricted to those reporting their wages in the CPS[48] and the dependent variable is the natural log of the gross weekly wage.[49] These characteristics ought to make the results more comparable with future studies.

5 Results

Industry and occupation do affect the magnitude of the wage-risk premium as well as the estimated coefficients on other covariates in a systematic manner. First, consider the covariates. The wage differences for gender and race (consistently significantly negative) get steadily smaller as more details on occupation are included. This makes sense that the more carefully defined the skill set is, the closer the wages, all else equal. The differential for education also decreases with more occupation controls, except for a slight rise when 21 controls are used. For all three of these regressors, the coefficient estimates decrease with the minimal number of industry controls (seven) but then are stable with more industry controls added. The coefficients on these are smallest in the presence of both the maximum occupation and industry controls, that is, the disaggregated set. There is also a clear pattern, but a different one, for the estimate of the wage-risk premium.

47 Burnham and Anderson (2004) propose pooling the information from multiple model variations. In this study, the AIC and BIC are used only to compare the effects of the various combinations of industry and occupation controls.

48 Restricting the sample to reported, rather than imputed, wages is discussed in note 29.

49 The weekly wage as dependent variable is likely to be a more consistent estimate than hourly wage for a sample of both salary and hourly workers. Studies using hourly wages from the CPS data convert weekly wages for those earning a salary into hourly wages by dividing the weekly wage by the usual number of hours worked per week. For salaried workers, the usual hours reported is difficult to interpret and over 40% of the sample are paid a salary rather than an hourly wage. On the other hand, the gross weekly wage for hourly workers is hourly wages times the reported usual hours. I follow the recommendation of Thaler and Rosen (1976) and use the weekly wage.

Table 3 presents the coefficient estimates on risk* for four of the models.[50] Panel A compares the square and long matrices for risk measures constructed from data on all workers using the CPS as the denominator source. Panel B compares measures based on data for wage and salary workers only, using the CPS as the denominator source. Panel C also reports on measures constructed from data on wage and salary workers only now with the denominator from the OES.

There is a pattern of dramatic differences in the estimates of β, the wage-risk premium, due to model variations (reading across) with only slight differences due to how risk is measured (reading down). When neither industry nor occupational controls are included, (Model 1, the first column in each panel), the estimated coefficients for all six measures are similar in magnitude and are statistically significant ($p<0.05$). In the second column, Model 2 (73 industry controls and no occupational controls), the coefficient is smaller (and sometimes <0) than the corresponding value from Model 1 regardless of how risk is measured and is not statistically significant. Whatever premium might be attributed to the difference in riskiness of the job appears to be absorbed into the differences in wages based solely on the industry. This is not unexpected, considering the literature on industry wage differentials discussed above.[51] The attenuation of the coefficient on risk* in the presence of industry controls is found even with few industry controls (beginning with seven, highest level of aggregation). This result indicates that without including industry fixed effects in the wage equation, the estimate for the coefficient on risk* includes wage effects due to inter-industry differences.

Occupation controls have an opposite effect. See the third column for Model 3, with 21 occupation and no industry controls, where for every risk* measure, the coefficient is larger, and the standard error smaller, compared to the no control

50 The results for all models for each of the 6 measures can be found at <http://faculty.knox.edu/cscotton/expcover.pdf>. For all models, regressors have the expected sign and are stable with respect to risk*. The results for the model used in the comparison described on Table 2, with no industry controls and nine occupational controls, are included there. The coefficients, however, are not comparable with those on Table 2 because the dependent variables and sample are different. Models with hour wage, or ln (hour wage), as the dependent variable and the sample restricted to those reporting wages only do yield the same pattern of results. Consistent with Bollinger (2001) and Hirsh (2008), the coefficient estimates on death risk are larger in the restricted sample while most of the other regressors are slightly smaller.

51 Not reported on the table but worth noting is that all but eight of the 72 industry controls are statistically different from zero ($p<0.1$ or less) when the risk measure is from the square matrix. This is robust across all risk* and regardless of the number of occupational controls. When the risk measure is from the long matrix – defined by more detailed industries – only as many as 10 industries may be no different from zero. The excluded industry is oil and gas extraction. Similar results are found when no risk measure is included. The magnitude, and the level of significance, of the inter-industry wage differential is influenced by the use of occupational controls.

Table 3 Coefficient estimates on risk* [dependent variable = ln (gross weekly earnings); LF sample n=84,336].

	Model			
	1	2	3	4
	No industry or occupation controls	73 industry and no occupation controls	No industry and 21 occupation controls	73 industry and 21 occupation controls
Panel A: Fatal risk construction includes all workers and CPS estimates for denominator.				
Square_All_CPS	0.0555**	-0.0172	0.0929***	0.0393***
(se)	(0.0232)	(0.0149)	(0.0125)	(0.0113)
R^2	0.420	0.489	0.499	0.529
AIC	93,144.57	82,492.03	80,698.37	75,560.80
BIC	93,303.39	82,650.85	80,857.19	75,906.48
Long_All_CPS	0.0526***	−0.0076	0.0799***	0.0407***
(se)	(0.0165)	(0.0110)	(0.0098)	(0.0079)
R^2	0.420	0.489	0.468	0.529
AIC	93,127.86	82,508.42	80,695.25	75,534.80
BIC	93,286.68	82,667.24	80,854.07	75,880.48
Panel B: Fatal risk construction includes only WS workers with CPS denominator.				
Square_Wage_CPS	0.0574**	−0.0232	0.1015***	0.0443***
(se)	(0.0232)	(0.0157)	(0.0119)	(0.0111)
R^2	0.420	0.489	0.500	0.529
AIC	93,116.64	82,475.25	80,605.58	75,548.83
BIC	93,275.47	82,634.07	80,764.40	75,894.51
Long_Wage_CPS	0.0543***	−0.0120	0.0861***	0.0456***
(se)	(0.0165)	(0.0119)	(0.0094)	(0.0078)
R^2	0.420	0.489	0.500	0.529
AIC	93,095.78	82,501.59	80,607.69	75,518.94
BIC	93,254.60	82,660.41	80,766.51	75,864.61
Panel C: Fatal risk construction includes only WS workers with OES denominator.				
Square_Wage_OES	0.0588***	−0.0018	0.0774***	0.0336***
(se)	(0.0214)	(0.0124)	(0.0137)	(0.0106)
R^2	0.422	0.488	0.500	0.529
AIC	92,873.01	82,512.71	80,630.91	75,550.75
BIC	93,031.83	82,671.53	80,789.74	75,896.43
Long_Wage_OES	0.0551***	0.0038	0.0689***	0.0371***
(se)	(0.0151)	(0.0098)	(0.0081)	(0.0065)
R^2	0.422	0.488	0.500	0.529
AIC	92,845.31	82,511.49	80,599.03	75,518.48
BIC	93,004.14	82,670.32	80,757.85	75,864.15

Robust standard errors in parentheses based on clusters: for square matrix = 330; for long matrix = 597. ***p<0.01, **p<0.05, *p<0.1.

case of Model 1.[52] Indeed, using any set of occupational controls in combination with any industrial controls increases the estimate of β. When both the least aggregated industry (73) and occupation (21) controls are employed as shown in the last column, Model 4, the coefficient estimates are smaller than when no controls are used.[53] Not shown is the model using the 19 industry and 21 occupation controls (similar to the aggregation in the square matrix) or the one using 73 industry controls and nine occupation controls (similar to the aggregation of the long matrix) since these not appreciably different from Model 4.

The coefficients are smaller yet, but remain statistically significant (at least $p<0.1$), when industry and occupation controls are at the most detailed level available in the labor force data as reported in Table 4. The strong consistent patterns demonstrated here illustrate the sensitivity of the magnitude of the wage-risk premium to the presence of industry and occupational controls while stable depending upon how the measure is constructed. This suggests the fatal risk measure, based on industry *and* occupation, no matter how constructed, is robust when controlling for *both* the industry and occupation fixed effects.

Similar to the finding reported on Table 2, the R^2 changes little depending upon risk*; and it varies as expected by the number of industry and occupational controls used. While R^2 necessarily reflects the number of explanatory variables, the AIC and BIC values measure how well the additional variables fit. "Given two models fit on the same data, the model with the smaller value of the information criterion is considered to be better" (STATA base reference PDF, p. 158). These measures, reported on Table 3, decrease as more controls are used. They are smallest when the maximum industry and occupation controls are used as shown on Table 4. Both the AIC and BIC also indicate a better fit, a consistently lower value, when risk* is constructed based on wage earners only. The information criteria slightly favor those models using risk measured by the long matrix of jobs.

The standard errors of the coefficients show a similar pattern. When risk is measured by the long matrix of jobs, the standard error on risk* is smaller than with the square matrix of jobs. However, with no controls, or with occupational controls alone, the coefficient is greater in magnitude. Yet, this pattern goes away when both controls are employed.

52 With or without a fatal risk measure in Model 3 (with no industry controls, 21 occupational controls), 18 of the 20 occupational controls are statistically different from zero ($p<0.01$). There is no appreciable effect on the statistical significance of the occupational controls themselves as compared to the model with no fatal risk measure.

53 The pattern found with industry and occupational controls, individually or combined, holds with just a few categories of either. The estimates are much more sensitive to few controls, suggesting that industry differentials are very predominate and specific to how the industries are aggregated.

Table 4 Coefficient estimates on risk* with 240 industry and 480 occupation controls.

	Square_ All_CPS	Square_ Wage_CPS	Square_ Wage_OES	Long_ All_CPS	Long_ Wage_CPS	Long_ Wage_OES
Risk*	0.0147*	0.0161**	0.0209***	0.0176***	0.0202***	0.0217***
(se)	(0.0079)	(0.0079)	(0.0063)	(0.0067)	(0.0067)	(0.0049)
R^2	0.578	0.578	0.578	0.578	0.578	0.579
AIC	66,684.2	66,683.3	66,671.68	66,677.97	66,674.71	66,660.49
BIC	69,066.55	69,065.65	69,054.04	69,060.32	69,057.07	69,042.84

Robust standard errors in parentheses; *p<0.1, **p<0.05, ***p<0.01.

6 Discussion

The intent of this study is not to posit a particular value for the VSL; rather, it is to demonstrate how the construction of the fatal risk rate measure impacts the magnitude of the VSL estimate. Clearly, different constructions of the fatal risk rate produce different estimates of the wage-risk premium, all else equal, as demonstrated on Table 2. Further, the rows in Table 3 indicate that the same construction of the risk measure will yield a widely varying estimate of the wage-risk premium depending only on differences in industry and occupational controls. The industry and occupation controls work together, but differently, in influencing the magnitude and precision of the estimate. Regardless of decisions in the construction of the measure, the fixed effects of the worker's industry and occupation do determine the size of the wage-risk premium. The fact that industry controls can be included in the wage equation along with occupational controls resolves concerns about the fatal-risk rate based on industry and occupation being able to "stand up" to inter-industry differentials. Since industry or occupation alone, regardless of risk, significantly determine the wage clearly more, rather than fewer, controls for industry *and* occupation ought to be employed when estimating a wage-risk premium. The least aggregated of the established hierarchies would seem to be the most appropriate choice to assure consistency across studies.

It is clear, too, that construction choices do matter. Measures including all workers in the risk rate produce estimates slightly smaller than measures based on wage and salary workers only.[54] The measures constructed based on wage

54 See Table 4 with the maximum number of controls. The results are mixed depending upon the denominator and as the number of controls vary as shown in Table 3 Panels B and C compared to Panel A.

earners only have a greater variability than those for all workers, signifying greater differences in the relative riskiness of the jobs in the payroll labor market (see Appendix 2). In general, self-employed workers face different risks and have different work environments (Pegula, 2004). This may be attributable to differing attitudes toward following safety standards and in taking risks or because of inherent dangers of working in isolation. In the case of automobile mechanics for example, (regardless of industry) approximately 16% are self-employed but they account for over 30% of automobile mechanic fatalities (Smith, 2007).[55] While overall deaths are disproportionally high among the self-employed, their work environment in some respects may be considered safer. For example, self-employed electricians, and other skilled tradesmen, are more likely to be on smaller worksites without the greater risks faced by electricians working at industrial construction sites.

These cases suggest other factors impacting risk which are wholly distinct from the specific assignment into a job cluster. If such differences are not random, but systematic, lumping the self-employed with the wage earners to produce a single measure for on-the-job risk would bias the wage-risk premium estimate. These differences are likely to be magnified when methods are employed to address issues such as endogeneity or heterogeneity of risk, which are not dealt with in this study.

Furthermore, including the self-employed in measures of relative risk invites a higher likelihood for error in the data itself. To be counted in the CFOI the injury must have happened while "on-duty". But, many of the self-employed work from home or are frequently in transit where work-related and personal chores are less definitively separable than is the case for wage earners. The additional human judgment required to determine whether a fatality is work related could be expected to introduce a possible systematic mis-measurement to the numerator. Finally, self-employed workers tend to wear a variety of different hats to accommodate their activities to the shifting needs of the market place. While coding in the CFOI does indicate the circumstances surrounding the fatal incident, this does not impact the coding for the worker's industry and occupation. Subsequent assignments into a particular job cluster will be arbitrary in some cases. To be consistent with theory, and because the data are available, the risk rates used in estimating a wage-risk premium ought to be derived from counts of deaths among wage earners only.

55 Indeed, how one dies is also an issue for consideration not dealt with in this study. Consider, for example that "nearly half (45.8%) of the self-inflicted fatalities [among mechanics] were by self-employed mechanics, although only 16% of mechanics are self-employed". (Smith 2007). See also Scotton and Taylor for an illustration of how the risk rate can be disaggregated to better reflect how to value the type of risk faced or being reduced.

Whether to use the square or long matrix (the configuration of the job clusters underlying the fatal risk measure) to construct the risk rate is not as clear. Comparing the risk rates assigned to workers in a labor force sample illustrates the variations between, or better said, the limitations of, these two matrices. Table 5 provides the risk rates for a few select jobs from both matrices. It shows a worker in the Transportation and Material Moving (TMM) occupation in the Construction industry is assigned the same rate (2.4209)[56] in both the square and the long matrix because there is no difference in the aggregation of this occupation/industry pair between the two matrices.[57] But there are many jobs where the rates will be very different for the *same* worker depending upon which matrix is used for risk rate assignment. While TMM occupations are aggregated in both matrices, the long matrix disaggregates most industry groups, while the square matrix does not. One example is the Hotel and Restaurant (H&R) industries. These are not disaggregated in the square matrix, so any TMM worker in H&R would be assigned the rate of 1.0142, regardless of the specific industry. The long matrix, though, distinguishes two separate sub industries within H&R. This means the TMM worker in the Hotel Accommodations industry would be assigned a rate of just 0.6198 in the long matrix, but a rate of 1.0142 when using the square.

Consider now the effect of disaggregating Service occupations in the square matrix, but not the long. Within the Construction industry, the risk in these occupations ranges from 0.8772 to 4.5083 when computed using the square matrix configuration, which distinguishes up to five service occupations. (Table 5 shows two of the Service occupations. The particular rates are not reported because of BLS confidentiality requirements.) But any Service worker in "Construction" would be assigned the rate of 1.6164 when using the long matrix where Service workers remain clustered into a single occupational group. In the long matrix, Service workers are certain to have their relative risk under or over estimated in an undetermined but non-random way.

The difference in rate assignment depending upon the matrix choice is more complex, for example, for Service workers in the H&R industry. In the square matrix, the risk rate varies (from 0 to 3.1433)[58] depending upon which of the five service occupations the worker is in but does not vary for any of the two sub industries. When the long matrix is used, regardless of one's Service occupation, the risk varies only by which of the two sub industries the worker is in. A few examples will illustrate. Protective service workers in the hotel accommodations industry (or any H&R industry) are assigned the rate of 3.1333 in the square

56 All risk rates are per 10,000 workers.

57 Appendix 1 shows the industries and occupations in each matrix.

58 At least one of the service occupations has zero risk in this industry group.

Table 5 Comparison of risk rates[a] from square and long matrix for selected job clusters.

	Industries							
	Construction		Hotel and restaurant		Hotel accommodations		Restaurants and bars	
Occupations	**Square**	**Long**	**Square**	**Long**	**Square**	**Long**	**Square**	**Long**
Transportation and Material Moving	2.4209	2.4209	1.0142	0.6198–1.0546	1.0142	0.6198	1.0142	1.0546
Service	0.8772–4.5083	1.6164	0–3.1433	0.1171–0.1419				
Protective Services			3.1433		3.1433	0.1419	3.1433	0.1171
Food Prep and Serving			0.0915		0.0915	0.1419	0.0915	0.1171

[a]Risk measure is deaths per 10,000 workers, based on all workers (defined in Figure 1 as Square_All_CPS or Long_All_CPS). All fatal injury rates were generated by the author with restricted access to the BLS CFOI microdata.

matrix. While that worker (or any Service occupation worker in "Hotel Accommodations") is assigned the rate of 0.1419 in the long matrix. As the Table shows, by using the long matrix it appears working in "Hotel Accommodations" is relatively more risky for service occupations than working in "Restaurants & Bars." On the other hand, those in Protective Service occupations face relatively more risk, in the H&R industry as a whole, than Food Prep workers if risk is assigned by the square matrix, but the long matrix assigns them the same risk.

Actually, none of the measures shown in Table 5 (created by either the square or long matrix) represent the "true risk" of the Protective Service worker in "Hotel Accommodations" versus "Restaurants & Bars." The rate is either for all service workers in the sub industry (long matrix) or for all Protective Service workers in both sub industries (square matrix). This leaves the question of whether the expected wage-risk premium is based on the difference between working at a hotel versus a restaurant or between working as a bouncer versus waiting tables. In the case of the TMM worker, the long matrix does differentiate the risk between working in "Hotels" versus in "Restaurants" only because there is no disaggregation for the particular occupations within this category.[59] (Among other possible distinctions at the 3-digit SOC for TMM one could isolate drivers, for instance, from all other TMM workers.) These examples illustrate how the matrix construction changes the relative risk for a particular worker (observation) in the labor force data set used to estimate the wage-risk premium, while at the same time they indicate how neither the square nor long matrix can be said to be the better or the correct matrix for our purposes in VSL estimation. Both have strengths; but for people in most jobs, neither represents the true (objective) relative risk.

The two configurations do, however, serve the purpose of illustrating that the matrix choice does matter in the estimation of a wage-risk premium because determining how to configure the job matrix determines how risk is assigned. Clearly, either of the two matrices, following a "canned" hierarchy, provides a better set of relative risk rates than available previously when rates were created based only on industry or only by occupation. So, we have solved the secretary and coal miner problem. Now, since the data are available, a better matrix could be developed, one that drills down to an appropriate level of detail by both industry and occupation to capture significant relative risk across jobs. It may not be relevant, practical, or possible[60] to disaggregate every occupation group and every industry group, but

59 It would be possible to disaggregate TMM occupations up to seven sub groups at the 3-digit SOC level.

60 With any disaggregation beyond the square matrix, the matching and merging of data needed for D_{oi} and W_{oi} becomes more problematic. And, the rate produced must be able to be matched to a labor force data set.

a systematic development of relevant clusters for VSL estimation overriding the default hierarchies in the coding books now needs to be addressed.

A final factor related to the construction of the measure illuminated by this study, but not fully resolved, is the source of the denominator. The effects when using the CPS versus the OES for estimates of the employment level when risk rates are constructed from counts among wage earners only can be directly compared. (Compare *wage_CPS* to *wage_OES* on Tables 3 and 4.) Estimates for the wage-risk premium are more sensitive to the presence of greater industry and occupation controls using the CPS estimates and the AIC and BIC are smallest using the OES estimates with the greatest number of controls. These results favor the OES. The OES would also enable greater disaggregation in industries such as "Construction," which the CPS does not.[61] In addition, the OES estimates, coming from firms (rather than the CPS household survey) may be more accurate, or at least more consistent, for estimates of employment levels for wage earners. On the other hand, OES estimates may not reflect employment well in firms with few employees, resulting in under counts of employment levels. In all, a better case can be made for using the OES as a source for employment levels. Problems remain, however, with any choice for the denominator. Further study of the effects of various options (including, for example, using hours worked) is needed.

What this study illustrates clearly is that the greater impact on the risk premium results from varying the fixed effects of the workers' industry and occupation rather than the construction of risk measure itself, at least with the two matrices used here. Industry controls, whether risk is based on jobs defined by the square or long matrix, in effect attenuate the wage-risk premium. Including occupation controls, on the other hand, increases the magnitude of the estimate as well as reduces the standard error. Smaller standard errors on the risk coefficient result when both industry and occupational controls are used in the model. While the problem with the matrix choice appears to be ameliorated when controlling for both industry and occupation in the estimation, it is still worthwhile to create a more appropriate job matrix to improve the measure of the relative risk.

7 Conclusion

Two critical conclusions can be drawn from these findings which should figure prominently in estimating the VSL. First, using fatal workplace risk rates when

61 If the CPS is used for the labor force sample where workers are not distinguished by type of Construction industry this would have no advantage.

based on a worker's occupation within a particular industry group, and also controlling for the worker's occupation and industry, do yield a positive and statistically significant wage-risk premium. This confirms long standing theory on compensating wage differentials and shows that an objective measure of the probability of death on the job plays a role in determining wages independent of other differences in wages due to occupation or industry. Most importantly, models not controlling for both industry *and* occupation, regardless of the risk measured used, do bias the estimate of the wage-risk premium, resulting in an arbitrary estimate of the VSL.

Secondly, the magnitude of wage-risk premium is highly sensitive to the particular level of aggregation used in controlling for industry and occupation. This sensitivity, in combination with decisions made in the construction of the measure, can account for estimates of the wage-risk premium ranging between 2% and 10% of gross weekly wages (as shown on Table 3), corresponding to a range in the VSL estimate from less than $8 million to more than $40 million.

While this paper addresses some key modeling issues, there are other modeling decisions (and many remaining criticisms) to be addressed. This study shows that using a VSL estimate from any particular labor market study is problematic given the range demonstrated here under controlled conditions. This range in the estimates of the wage-risk premium also highlights the futility of combining results from diverse studies without taking the construction details of the measure, as well as modeling and sample factors, into consideration. These unknown inputs could skew the estimate in unknown directions. Unless studies use common metrics for the measure, estimates from various studies will remain difficult to compare let alone combine. The results from this study affirm the Science Advisory Board's recommendation to the EPA for greater consistency and transparency in studies estimating a VSL.

Thaler and Rosen's first efforts to estimate a VSL were based on actuarial data of differences in death rates depending upon one's occupation but unrelated to the workplace. Soon after, data on workplace deaths by industry alone were used; later, rates were also developed based on occupation only. These earlier estimates were rightly criticized for not capturing the relative riskiness of particular jobs. Now that attention has become focused on constructing measures for occupations within industries a new set of concerns arises. The necessary choices identified, tested and compared within this study had heretofore not been explicitly declared and thus were not accessible for review. The richness of the CFOI data, along with clear evidence that an objective measure of fatal occupational risk is robust in the presence of industry and occupational controls, enables a more theoretically consistent measure to be constructed and used in the estimation of the VSL. Developing standard criteria for a sound measure of workplace risk, not subject to private, possibly arbitrary (or unrecognized) choices, remains a necessary step to convince decision makers of the legitimacy of using labor market data to estimate specific values of the costs and benefits of reducing risk.

Acknowledgement: This research was conducted with restricted access to Bureau of Labor Statistics (BLS) data. The views expressed here do not necessarily reflect the views of the BLS. The project benefited by Professor Richard Stout's consultation and critical insights. Discussants and conference participants at the Southern Economic Association annual meeting, Western Economic Association International Conference, and Midwest Economics Association Conference commented on early drafts. Sam Flanders, Sean Frohling, Boram Lee, Steven Major, and Sharon Clayton provided assistance with the research. Two anonymous referees provided valuable comments which improved this paper.

Appendix 1

Occupational groups.

In the square matrix (n=22)	Number of industries within the occupation	In the long matrix (n=10)	Number of industries within the occupation
Management	20	Management, Business and	76
Business and Finance	20	Finance	
Computer and Math	20	Professional and Related	76
Architecture and Engineering	20		
Life, Physical and Social Science	20		
Community and Social Service	14		
Legal Work	17		
Education and Library	18		
Arts, Entertainment, Media	20		
Healthcare Providers	20		
Healthcare Support	15	Service	75
Protective Services	20		
Food Prep and Serving	19		
Building and Grounds Maintenance	20		
Personal Care and Service	18		
Sales and Related	20	Sales and Related	76

(**Appendix 1** Continued)

In the square matrix (n=22)	**Number of industries within the occupation**	**In the long matrix (n=10)**	**Number of industries within the occupation**
Office and Administration Support	20	Office and Administration Support	76
Farming, Fishing and Forestry	18	Farming, Fishing and Forestry	42
Construction and Extraction	20	Construction and Extraction	68
Installation, Maintenance and Repair	20	Installation, Maintenance and Repair	76
Production	20	Production	74
Transportation and Material Moving	20	Transportation and Material Moving	75
Total	419	Total	714

Industry groups.

In the square matrix (n=20)	**Number of occupations within the industry**	**In the long matrix (n=76)**	**Number of occupations within the industry**
Agriculture, Forestry, Fishing and Hunting	19	Forestry and Logging	10
		Support Activities for Agriculture and Forestry	10
Mining	18	Oil and Gas Extraction	9
		Mining (except oil and gas)	10
		Support Activities for Mining	10
Utilities	18	Utilities	10
Construction	20	Construction	10
Manufacturing	21	Food Manuf	10
		Beverage and Tobacco Product Manuf	10
		Textile Mills/Apparel Manuf	9
		Textile Product Mills	9
		Leather and Allied Product Manuf	8
		Wood Product Manuf	10
		Paper Manuf	10
		Printing and Related Support Activities	9
		Petroleum and Coal Products Manuf	9
		Chemical Manuf	10
		Plastic and Rubber Products Manuf	9

(**Appendix 1** Continued)

In the square matrix (n=20)	Number of occupations within the industry	In the long matrix (n=76)	Number of occupations within the industry
		Nonmetallic Mineral Manuf	10
		Primary Metal Manuf	9
		Fabricated Metal Manuf	9
		Machinery Manuf	10
		Computer and Electronics Manuf	9
		Electrical Equip, Appliance and Component	9
		Transportation Equip Manuf	9
		Furniture Related Manuf	10
		Miscellaneous Manuf	10
Wholesale Trade	22	Durable Goods Wholesalers	10
		Nondurable Goods Whole-salers	10
		Wholesale Electronic Mrkts	9
Retail Trade	22	Motor Veh and Parts Dealers	9
		Furniture and Home Furnishing	9
		Electronic and Appliance Stores, Sporting	10
		Building Material and Garden Equip Dealer	10
		Food and Bev Stores	10
		Health and Personal Care Stores	9
		Gas Stations	9
		Clothing and Accessories Stores	8
		General Merchandise Stores	10
		Miscellaneous Store Retailers	10
		Non-store Retailers	9
Transportation	21	Air transport	8
		Rail Transport	9
		Water Transport	9
		Truck Transport	10
		Transit and Ground Passen-ger Transport	9
		Pipeline Transport	9
		Scenic and Sightseeing Transport	9
		Transport Support	9

(**Appendix 1** Continued)

In the square matrix (n=20)	Number of occupations within the industry	In the long matrix (n=76)	Number of occupations within the industry
Couriers and Warehousing, except Postal	20	Couriers and Messengers	8
		Warehouse and Storage	10
Information	21	Publishing, Except Internet	9
		Movie and Recording	9
		Broadcasting and Telecoms	9
		Internet Broadcast and Publish	5
		Internet Service Providers and Search Portals	8
		Other Info Services	8
Finance and Insurance	22	Finance and Insurance	10
Real Estate	22	Real Estate and Leasing	10
Professional and Technical Service	22	Professional Scientific Technical Services	10
Management	20	Management of Enterprises	9
Admin Support, Waste Management	22	Admin and Support Services	10
		Waste Manag and Remediation Services	10
Education	22	Education	10
Healthcare and Social Services	22	Ambulatory Health Services	9
		Hospitals	9
		Nursing and Resid Care Facilities	10
		Social Services	10
Arts, Entertainment and Recreation	22	Performing Arts and Spectator Sports	10
		Museums and Historic Sites	10
		Amusement and Gambling	10
Hotel and Restaurant	21	Hotel Accommodation	10
		Restaurants and Bars	10
Other Services, except Public Administration and Private Households	22	Repair and Maintenance	10
		Personal and Laundry Services	10
		Religious, Grant Making, Civic and Professional Organizations	10
Total	419	Total	714

Appendix 2[62]

Distribution of risk*.

	Square_ All_CPS	Square_ Wage_CPS	Square_ Wage_OES	Long _ All_CPS	Long _ Wage_CPS	Long _ Wage_OES
Panel A: Weighted distribution of risk rates, fatalities per 10,000 workers.						
Mean	0.3541	0.3365	0.3995	0.3556	0.3388	0.4022
Robust estimate of mean	0.1165	0.0969	0.1225	0.0999	0.0843	0.1010
p5	0.0239	0.0222	0.0200	0.0220	0.0205	0.0192
p10	0.0300	0.0256	0.0246	0.0369	0.0336	0.0289
p25	0.0459	0.0449	0.0486	0.0449	0.0427	0.0475
p50	0.1138	0.0978	0.1437	0.1128	0.0918	0.1079
p75	0.2581	0.2411	0.2697	0.3221	0.2450	0.3293
p90	1.2330	1.1949	1.1456	1.2642	1.2535	1.3976
p99	2.8754	2.8593	3.5445	3.4277	3.5785	4.4357
Max	4.7712	4.6974	13.3971	9.2138	9.3637	7.7207
Panel B: n=84,336, number of workers in LF sample. Number (%) of workers in the sample with:						
Rate=0	1365 (2%)	1592 (2%)	1592 (2%)	1684 (2%)	1739 (2%)	1739 (2%)
Rate>0 & ≤0.25	60,840 (72%)	63,322 (75%)	59,049 (70%)	58,793 (70%)	62,225 (74%)	57,422 (68%)
Rate >0.25 & ≤0.50	5545 (7%)	3464(4%)	5055 (6%)	8331(10%)	5139(6%)	8283 (10%)
Rate >0.50 & ≤1.0	7008 (8%)	6741 (8%)	9575 (11%	5301 (6%)	5318 (6%)	7132 (8%)
Rate > 1	9578 (11%)	9217 (11%)	9065 (11%)	10,227 (12%)	9915 (12%)	9760 (12%)
Panel C: Un-weighted distribution of risk rates, fatalities per 10,000 workers.						
n=	419	419	419	714	714	714
mean	0.4863	0.4547	1.1187	0.6656	0.6214	1.5673
Robust estimate of mean	0.0941	0.0858	0.0989	0.1857	0.1474	0.2086
p25	0	0	0	0.0282	0.0149	0.0134
p50	0.0973	0.0941	0.1085	0.1713	0.1389	0.1903
p75	0.4139	0.3901	0.5577	0.6332	0.6193	0.7193
p90	1.0696	1.0000	1.5198	1.5286	1.4813	2.0607
p99	4.7712	4.6974	19.0840	9.2138	7.9612	13.2979
Max	17.0068	17.0068	125.0000	17.3611	17.3611	375.0000

62 All fatal injury rates were generated by the author with restricted access to the BLS CFOI microdata.

(**Appendix 2** Continued)

	Square_ All_CPS	Square_ Wage_CPS	Square_ Wage_OES	Long _ All_CPS	Long _ Wage_CPS	Long _ Wage_OES
Panel D: n=number of jobs. Number (%) of jobs with:						
Rate =0	131 (31%)	138 (33%)	138 (33%)	168 (24%)	176 (25%)	176 (25%)
Rate >0 & ≤0.25	147 (35%)	148 (35%)	126 (30%)	239 (33%)	255 (36%)	215 (30%)
Rate >0.25 & ≤0.50	46 (11%)	42 (10%)	41 (10%)	97 (14%)	82 (11%)	90 (13%)
Rate >0.50 & ≤1.0	47 (11%)	50 (12%)	53 (13%)	89 (12%)	89 (12%)	103 (18%)
Rate >1	48 (11%)	41 (10%)	61 (15%)	121 (17%)	112 (16%)	130 (18%)

References

Ashenfelter, O. (2006). *Measuring the value of a statistical life: problems and prospects*. [Discussion Paper No. 1911]. Bonn, Germany: IZA.

Ashenfelter, O., & Greenstone, M. (2004). Estimating the value of a statistical life: the importance of omitted variables and publication bias. *The American Economic Review 94,* 454–460.

Black, D. A., & Kniesner, T. J. (2003). On the measurement of job risk in hedonic wage models. *Journal of Risk and Uncertainty 27,* 205–220. DOI: 10.1023/A:1025889125822.

Bollinger, C. R. (2001). Response error and the union wage differential. *Southern Economic Journal 68*, 60.

Bollinger, C. R., & Hirsch, B. T. (2005). *Match bias from earnings imputation in the current population survey: the case of imperfect matching*. [Discussion Paper No. 1846]. Bonn, Germany: IZA.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research 33*, 261–304. DOI: 10.1177/0049124104268644.

Cameron, T. A. (2010). Euthanizing the value of a statistical life. *Review of Environmental Economics and Policy 4*, 161–178. DOI: 10.1093/reep/req010.

Colbert Report, The word segment (priceless). July 14, 2008. Available at: http://www.colbertnation.com/the-colbert-report-videos/176175/july-14-2008/the-word---priceless?videoId=176175. Accessed on November 14, 2008.

Costa, D. L., & Kahn, M. E. (2002). *Changes in the value of life, 1940–1980*. [Working Paper 9396]. Cambridge, MA: National Bureau of Economic Research.

Cropper, M., & Morgan, G. M. (2007). SAB advisory on EPA's issues in valuing mortality risk reduction. [Letter to EPA Administrator Stephen L. Johnson. Doc no. EPA-SAB-08-001]. Washington, DC: EPA.

Dickens, W. T., & Katz, L. F. (1987). Inter-Industry wage differentials and industry characteristics. In: K. Lang & J. S. Leonard (Eds.), *Unemployment and the Structure of Labor Markets* (pp. 48–89). New York: Basil Blackwell.

Drudi, D. (1995). The evolution of occupational fatality statistics in the United States. *Compensation and Working Conditions 47*, 1–5.

Fairris, D., & Jonasson, E. (2008). What accounts for intra-industry wage differentials? Results from a survey of establishments. *Journal of Economic Issues 42,* 97–113.

Gittleman, M., & Pierce, B. (2011). Inter-industry wage differentials, job content and unobserved ability. *Industrial and Labor relations review 64*, 356–374.

Gittleman, M., & Pierce, B. (2006). A different approach to measuring workplace safety: Injuries and fatalities relative to output. *Compensation and Working Conditions*. Retrieved from http://www.bls.gov/opub/cwc/sh20060724ar01p1.htm.

Goux, D., & Maurin, E. (1999). Persistence of interindustry wage differentials: a reexamination using matched worker-firm panel data. *Journal of Labor Economics 17*, 492–533. doi: 10.1086/209929.

Hammitt, J. K., & Robinson, L. A. (2011). The income elasticity of the value per statistical life: Transferring estimates between high and low income populations. *Journal of Benefit-Cost Analysis* 2, 1. DOI:10.2202/2152-2812 *2*, 1.

Hirsch, B. T. (2008). *Wage gaps large and small*. [Discussion Paper No. 3375] Bonn, Germany: IZA.

Hirsch. B. T., & Schumacher, E. J. (2004). Match bias in wage gap estimates due to earnings imputation. *Journal of Labor Economics 22*, 689–722.

Jones-Lee, M. (1974). The value of changes in the probability of death or injury. *The Journal of Political Economy 82*, 835–849.

Kochi, I., & Taylor, L. O. (2011). Risk heterogeneity and the value of reducing fatal risks: further market-based evidence. *Journal of Benefit-Cost Analysis 2*. Article 1. DOI: 10.2202/2152-2812.1079.

Krueger, A. B., & Summers, L. H. (1987). Reflections on the inter-industry wage structure. In: K. Lang & J. S. Leonard (Eds.), *Unemployment and the Structure of Labor Markets* (pp. 17–47). New York: Basil Blackwell.

Lane, J. I., Salmon, L. A., & Spletzer, J. R. (2007). Establishment wage differentials. *Monthly Labor Review 130*, 3–17.

Leigh, P. (1995). Compensating wages, value of a statistical life and inter-industry differentials. *Journal of Environmental Economics and Management 28*, 83–87. DOI: 10.1006/jeem.1995.1006.

Mathiowetz, N. A. (1992). Errors in reports of occupations. *Public Opinion Quarterly 56*, 352–355. DOI: 10.1086/269327.

Mellow, W., & Sider, H. (1983). Accuracy of response in labor market surveys: Evidence and implications. *Journal of Labor Economics 1*, 331–344. DOI: 10.1086/298016.

Moore, M. J., & Viscusi, W.K. (1988). Doubling the estimated value of a statistical life: Results using new occupational fatality data. *Journal of Policy Analysis and Management 73*, 476–490.

Mrozek, J. R., & Taylor, L.O. (2002). What determines the value of life? A meta-analysis. *Journal of Policy Analysis and Management 21*, 253–270. DOI: 10.1002/pam.10026.

National Bureau of Economic Research. (2007). Merged outgoing rotation group, *CPS Labor Extracts 1979-2006*. Cambridge, MA: NBER.

Northwood, J. (2010). Change to hours-based fatality rates in the census of fatal occupational injuries. *Compensation and Working Conditions*. Retrieved from http://www.bls.gov/opub/cwc/sh20100121ar01p1.htm

Osburn, J. (2000). Interindustry wage differentials: patterns and possible sources. *Monthly Labor Review 123,* 34.

Pegula, S. M. (2004) Occupational fatalities: self-employed workers and wage and salary workers. *Monthly Labor Review 127*, 30–40.

Personick, M. E., & Windau, J. A. (1995). Self-employed individuals fatally injured at work. *Monthly Labor Review 118*, 24–30.

Robinson, L. A. (2007) How US government agencies value mortality risk reduction. *Review of Environmental Economics and Policy* 1, 283–299. DOI: 10.1093/reep/rem018.

Robinson, L. A., Hammitt, J. K., Aldy, J.E., Krupnick, A., & Baxter, J. (2010). Valuing the risk of death from terrorist attacks. *Journal of Homeland Security and Emergency Management 7*, 14.

Ruser, J. W. (1998). Denominator choice in the calculation of workplace fatality rates. *American Journal of Industrial Medicine 33*, 151–156. DOI: 10.1002/(SICI)1097-0274(199802)33:2<151::AID-AJIM6>3.3.CO;2-3

Scharff, R. L. (2010). *Health-related costs from foodborne illness in the United States.* Washington, DC: Georgetown University, Produce Safety Project. Retrieved from Produce Safety Project http://www.producesafetyproject.org/admin/assets/files/Health-Related-Foodborne-Illness-Costs-Report.pdf-1.pdf

Schelling, T. C. (1968) The life you save may be your own. In: S.B. Chase, Jr. (Eds.), *Problems in Public Expenditure Analysis*. (pp. 127–162). Washington DC: The Brookings Institute.

Scotton, C. R. (2000). *The impact of workplace–risk measurement on estimates of the value of a statistical life* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (UMI no. 9962366).

Scotton, C. R., & Taylor, L. O. (2011). Valuing risk reductions: incorporating risk heterogeneity into a revealed preference framework. *Resource and Energy Economics 33*, 381–397. DOI: 10.1016/j.reseneeco.2010.06.001.

Smith, R. S. (1979). Compensating wage differentials and public policy: a review. *Industrial and Labor Relations Review 32*, 339–352. DOI: 10.2307/2522263.

Smith, S. M. (2007). Occupational injuries, illnesses, and fatalities to automotive service technicians and mechanics, 2003 to 2005. *Compensation and Working Conditions.* Retrieved from http://www.bls.gov/opub/cwc/print/sh20070521ar01p1.html.

StataCorp. (2011) . Stata. (Release 12) [Computer software]. College Station, TX: StataCorp.

Thaler, R. H. (1989). Anomalies: Interindustry wage differentials. *The Journal of Economic Perspectives* 3, 181–193.

Thaler, R. H., & Rosen, S. (1976). The value of saving a life: evidence from the labor market. In: Terleckyj, N. E. (Eds.), *Household Production and Consumption*. New York, NY: National Bureau of Economic Research.

US Bureau of Labor Statistics. (2007a). *Census of fatal occupational injuries*. Washington DC: US Bureau of Labor Statistics.

US Bureau of Labor Statistics. U.S. Department of Labor, (2007b). *Occupational Employment and Wages, 2006.* [Bulletin USDL-07-0712]. Washington, DC: Government Printing Office.

US Bureau of Labor Statistics, (2007c). *Occupational Employment Statistics*. Retrieved from http:// www.bls.gov/oes/.

US Census Bureau. (2003–2006). *Current population survey Data Dictionary.* (Data extracted via Data Ferret). Washington DC: US Census Bureau.

US Environmental Protection Agency. (1999). *Cost Benefit Analysis of the Clean Air Act.* Washington, DC: EPA.

US Environmental Protection Agency. (2010). *Guidelines for Preparing Economic Analyses.* Washington, DC: EPA. Retrieved from http://yosemite.epa.gov/ee/epa/eed.nsf/pages/guidelines.html.

US Office of Management and Budget. (2007). *North American Industry Classification System Manual*. [CD-Rom]. Springfield, VA: National Technical Information Service.
US Office of Federal Statistical Policy and Standards. (2000). *Standard Occupational Classification Manual*. Springfield, VA: National Technical Information Service.
Viscusi, W. K. (2004). The value of life: estimates with risks by occupation and industry. *Economic Inquiry 42*, 29–47. DOI: 10.1093/ei/cbh042.
Viscusi, W. K., & Aldy, J. E. (2003). The value of a statistical life: A critical review of market estimates throughout the world. *Journal of Risk and Uncertainty 27*, 5–76. DOI: 10.1023/A:1025598106257.