



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.



Robustness of VSL Values from Contingent Valuation Surveys

Anna Alberini

Working Paper Series

Working Paper # 05-01
January, 2005



U.S. Environmental Protection Agency
National Center for Environmental Economics
1200 Pennsylvania Avenue, NW (MC 1809)
Washington, DC 20460
<http://www.epa.gov/economics>

Robustness of VSL Values from Contingent Valuation Surveys

Anna Alberini

Correspondence:

Anna Alberini
AREC, 2200 Symons Hall
University of Maryland
College Park, MD 20742
aalberini@arec.umd.edu

NCEE Working Paper Series

Working Paper # 05-01
January, 2005

DISCLAIMER

The views expressed in this paper are those of the author(s) and do not necessarily represent those of the U.S. Environmental Protection Agency. In addition, although the research described in this paper may have been funded entirely or in part by the U.S. Environmental Protection Agency, it has not been subjected to the Agency's required peer and policy review. No official Agency endorsement should be inferred.

ROBUSTNESS OF VSL VALUES FROM CONTINGENT VALUATION SURVEYS *

By

Anna Alberini
Department of Agricultural and Resource Economics
2200 Symons Hall
University of Maryland

Address for correspondence:

Anna Alberini
AREC, 2200 Symons Hall
University of Maryland
College Park, MD 20742
USA

Phone 001 301 405-1267
Fax 001 301 314-9091
e-mail: aalberini@arec.umd.edu

ABSTRACT. This paper examines factors that may influence the estimates of the Value of a Statistical Life obtained from contingent valuation surveys that elicit the willingness to pay (WTP) for mortality risk reductions. We examine the importance of distributional assumptions, the choice of the welfare statistics of interest, the procedure for computing them, outliers, undesirable response effects, and internal validity of the WTP responses. We illustrate the importance of these factors using dichotomous-choice and open-ended WTP data from four recent contingent valuation surveys.

Key words: contingent valuation, VSL, WTP, risk reductions, robustness, outliers, endogeneity.

Subject Area Classifications: Health, Valuation, Benefit Cost Analysis

* This research was conducted under Cooperative Agreement 015-29528 with the U.S. Environmental Protection Agency. I wish to thank Drs. Johannesson, Persson and Gerking for making their data available to me. I also wish to thank my project officer, Nathalie Simon of the US Environmental Protection Agency, for her guidance with this project, seminar series participants at the U.S. EPA National Center for Environmental Economics and session attendees at the annual EAERE meeting in Budapest, Joel Huber, Trudy Cameron, Melonie Sullivan and Kelly Maguire for their comments on this work.

ROBUSTNESS OF VSL VALUES FROM CONTINGENT VALUATION SURVEYS

By

Anna Alberini

1. Introduction.

The Value of a Statistical Life (VSL) is the rate at which people are prepared to trade off income for a reduction in their risk of dying. The VSL is a key input for computing the mortality benefits of environmental and safety policies that save lives. In recent retrospective analyses of the Clean Air Act and of the Clean Air Act Amendments, for example, the US Environmental Protection Agency has used a VSL of \$6.1 million in its base analyses (\$3.7 million in “alternate” analyses), and the resulting monetized mortality benefits account for over 80% of total benefits of these environmental statutes. Within the European Commission, DG Environment uses central VSL estimates of about 1.2 million euro, with adjustments for age and for the futurity of the risk.¹

The VSL figures for cost-benefit analysis purposes are typically derived using three possible methods: (i) compensating wage studies, (ii) consumer behavior studies, and (iii) contingent valuation surveys. Compensating wage studies use data from labor markets to infer how much workers have to be compensated to accept riskier jobs—or the sacrifice in income they would agree to in exchange for an improvement in their workplace safety. Viscusi and Aldy (2003) document over 60 compensating wage studies conducted in 10 countries, noting that in the US most of the labor market studies produce estimates of the VSL in the range of \$4-9 million.

Consumer behavior studies observe tradeoffs between time and risk, or money and risk, to place a value on mortality risk reductions. An early such study (Blomquist,

¹ See http://europa.eu.int/comm/environment/enveco/others/recommended_interim_values.pdf.

1979) observed whether or not individuals fasten their seatbelts when driving. The VSL is calculated as the value of the time required for buckling up, divided by the reduction in the risk of dying in a traffic accident afforded by the use of seatbelts. Atkinson and Halvorsen (1990) obtain an estimate of the VSL from the higher price of cars with more sophisticated safety equipment. The VSL figures from labor market or consumer studies are often transferred to the environmental policy context. Doing so implicitly assumes that the preferences of individuals for income and risk do not vary with the context.

In contingent valuation surveys, respondents are asked to report their willingness to pay (WTP) for a specified—and hypothetical—risk reduction. Contingent valuation studies have the potential to circumvent many of the shortcomings for which the other approaches are sometimes criticized. For example, they lend themselves to valuing risk reductions in many contexts, and are thus not limited to workplace risks. Rather than assuming that people know the exact magnitude of the risks they face, in a well-designed CV study respondents are educated about them, and the extent of the risk reduction is spelled out explicitly for them.

Last but not least, CV allows the researcher to survey directly the beneficiaries of any proposed risk-reduction measure. This is regarded as a particularly advantageous feature of the method, because when the risk reduction measure is an environmental program, these beneficiaries (e.g., the elderly) are likely to be very different than the population covered in compensating wage studies, and may have different preferences for income and risk.

Despite these advantages and the flexibility of the approach, much debate surrounds the estimates of VSL from contingent valuation surveys. This paper focuses on

two main difficulties associated with the VSL figures from CV studies. The first is that many recent high-quality CV surveys have elicited information about WTP using dichotomous-choice questions. Dichotomous-choice questions have been shown to be incentive compatible (Hoeft and Randall, 1987), and are generally thought to be easier to answer than open-ended questions. However, the researcher must rely on assumptions about the distribution of the underlying WTP in order to obtain estimates of mean and median WTP, and these in turn are typically sensitive to the upper tail of the distribution of WTP.

Second, it is, in general, difficult to value risk reductions. Respondents are not used to dealing with probabilities, especially when risks are very small, and the cognitive burden imposed upon them in the survey—or the failure to communicate risks to them in a meaningful way—may result in undesirable effects, including failure to distinguish between risk reductions of different sizes (Hammitt and Graham, 1999), confusion between absolute and relative risk reductions (Baron, 1997), protest responses, completely random answers to the payment questions, etc. (Carson, 2000). One possible approach for uncovering these problems is to test for the internal validity of the responses, i.e., to check that WTP depends on certain variables in the ways predicted by economic theory. Alternatively, one may try to fit models that explicitly seek to identify abnormal responses.

The purpose of this paper is to discuss the importance of distributional and modeling assumptions, and the effect of econometric misspecifications and of the presence of abnormal response patterns in the sample on the estimates of the VSL. We illustrate our robustness criteria and checks using the data from four recent CV surveys.

We examine robustness with respect to four possible criteria. The first three are specific to dichotomous-choice WTP responses. First, we examine how the VSL changes with (a) the distribution WTP is assumed to follow, (b) the welfare statistic of interest (e.g., median or mean), given the distribution of WTP, and (c) given the distribution and the welfare statistic, different procedures for calculating the latter.

Our second set of analyses is similar to the first, except that it focuses on how (a), (b) and (c) impact the estimated relationship between WTP and specific covariates. Third, we examine outliers and abnormal response patterns in dichotomous-choice CV surveys. Fourth, we look at internal validity checks.

The remainder of this paper is organized as follows. Section II provides a definition of VSL and outlines the robustness criteria discussed in this paper. Section III focuses on dichotomous choice CV responses, examining the robustness of VSL to the assumed distribution of WTP, the welfare statistics used, and the procedure used for computing such welfare statistics. Section IV focuses on outliers and response mechanisms that do not comply with the economic paradigm, such as yea-saying, nay-saying, and completely random responses. Section V examines how internal validity tests, such as scope tests, are affected by the possible endogeneity of risks and WTP, and section VI focuses on the relationship between WTP and income. Section VII provides concluding remarks.

II. Econometric Robustness of VSL Figures

A. Definition of VSL

The Value of a Statistical Life is the rate at which individuals are prepared to trade off income for risk reductions. In an expected utility framework, let $U(w)$ be the (state-dependent) utility associated with income w if the individual is alive, and $V(w)$ the utility of income if the individual is dead. If the probability of dying is p , expected utility is defined as $(1-p)U(w)+pV(w)$. This expression can be further simplified to $(1-p)U(w)$ if it is assumed that $V(w)=0$ (i.e., the utility of income is zero when one is dead). The VSL is the rate of substitution between income and risk that keeps expected utility unchanged, and is in this context equal to $\frac{dw}{dp} = \frac{U(w)}{(1-p)U'(w)}$.

The VSL is, therefore, a derivative, but in practice contingent valuation surveys ask people to report information about their willingness to pay (WTP) for a specified—and finite—reduction in their risk of dying, Δp . The VSL is estimated as $WTP/\Delta p$. Accordingly, in this paper the robustness of the VSL estimates and the robustness of WTP estimates are regarded as interchangeable.

B. Robustness Criteria

This paper examines robustness with respect to four criteria. Because many recent high-quality CV studies have deployed dichotomous-choice questions to elicit information about WTP, the first series of robustness checks refers to dichotomous-choice CV data. Specifically, we examine by how much WTP changes with (a) the distribution WTP is assumed to follow, (b) the welfare statistic of interest (e.g., median or

mean), given the distribution of WTP, and (c) given the distribution and the welfare statistic, different procedures for calculating the latter.

Our second set of analyses is similar to the first, except that it focuses on how (a), (b) and (c) impact the estimated relationship between WTP and specific covariates, such as the age of the respondent.

Next, we turn to the issue of whether it is possible to identify abnormal responses to dichotomous-choice WTP questions. To identify outliers, we first use a “reduced-form” approach based on a regression equation relating the response to the payment question to observable individual characteristics, checking if the inclusion of these observations in the sample affects appreciably the estimates of WTP, and, if so, by how much. In our next step, we seek to model explicitly abnormal response patterns. By abnormal response patterns, we mean answers to the payment questions that do not comply with the economic paradigm, such as responses motivated by “yea-saying” or “nay-saying” behaviors, or completely random responses.

Fourth, we focus on internal validity. We first work with a dataset where respondents subjectively assessed their baseline risk and were asked to value a given reduction in this initial risk. Economic theory posits that WTP should increase appreciably with the size of the risk reduction. Moreover, meaningful VSL figures can be computed only if individuals are valuing the absolute risk reduction. We examine whether our ability to empirically check that the WTP responses are consistent with economic theory and with the VSL construct is affected by treating WTP as econometrically endogenous with risk. The specific application we use to explore these

issues elicits WTP using open-ended questions, producing observations of WTP on a continuous scale.

Economic theory also predicts that WTP should increase with respondent's income. We check how the estimated income elasticity of WTP changes as respondents who would be willing to commit a large fraction of their income to the risk reduction are excluded from the sample. This has potentially important consequences for benefit transfer, i.e. the practice of applying the results of the study conducted at one locale to another population or context.

III. The Data.

We illustrate out robustness checks using data from four applications, which we summarize in table 1. In the first application (Johannesson et al., 1997), a representative sample of Swedish adults aged 18-74 were surveyed over the telephone about their WTP for a reduction in their risk of dying. The survey was conducted in November 1996, and produced WTP data for a total of 2029 individuals, for a response rate of 83 percent. The goal of the study was to study the relationship between the VSL and the age of an individual.

Respondents were told that X out of 10000 people of their gender and age would die during the next year.² They were also asked to assume that a preventive and painless treatment was available that would reduce by 2 in 10000 the risk of dying in the next year, but have no effects thereafter. Information about WTP was elicited using single-bounded dichotomous-choice questions, with bid values ranging from 300 to 10000

² The baseline risk of death over the next year was 10, 30, 70, and 200 for males in the age groups 18-39, 40-49, 50-59, and 60-69, respectively. For females, the baseline risk values were 5, 20, 40, and 100. All baseline risks are out of 10000.

SEK.³ Johanesson et al. estimate the mean WTP to be equal to 6300 SEK (about 954 US dollars), which corresponds to a VSL of 31.4 million SEK (4.75 million US dollars), and detect a quadratic relationship between age and WTP that peaks at 40 years of age.

The second and third applications used in this study employed dichotomous choice questions with follow-ups, and a virtually identical survey instrument for Hamilton, Ontario (Krupnick et al., 2002) in Spring 1999 and a national sample of US respondents (Alberini et al., 2004) in August 2000. Both questionnaires were self-administered by the respondent using the computer. In the Hamilton study, respondents were asked to go to a centralized facility to take the survey, whereas in the US studies they received the questionnaire via Web-TV™.

By asking people to value immediate and future risk reductions (for which payment would have to start immediately), and by recruiting individuals of various ages, including the elderly, and health statuses, these studies explore four main research questions. The first is the relationship between VSL and age. The second is the relationship between WTP and the health status of the respondent. This is important for policy purposes, as some agencies have argued in favor of using Quality Adjusted Life-Years (QALY), a construct widely used in medical decisionmaking where values are adjusted for quality of life, which is presumably lower for chronically ill people.

The third research question is whether the WTP for a future risk reduction is less than the WTP for an immediate risk reduction (as is implied by discounting and by the fact that the individual may die before he reaches the age when the future risk reduction

³ The payment question read as follows: “It is estimated that X(Y) men (women) out of 10,000 in the same age as you will die during the next year. Assume that you could participate in a preventive and painless treatment which would reduce the risk that you will die during the next year, but has no effects beyond that year. The treatment reduces the risk of your dying during the next year from X(Y) to X-2 (Y-2) out of 10,000. Would you at present choose to buy this treatment if it costs SEK I?”

would begin), and the fourth is how large is the implicit discount rate (Alberini et al., 2004).

The fourth study (Persson et al., 2001) is a mail survey eliciting WTP for reductions in the risk of dying in a road-traffic related fatality. The survey was conducted in Sweden in Spring 1998. Questionnaires were mailed to a representative sample of Swedes of ages 18-74, for a total of 2884 returned questionnaires (the response rate was 51%). Two versions of the questionnaire were created. The first focused on the risk of non-fatal injuries, while the other focused on the risk of dying in a road-related accident. In this paper, attention is restricted on the 935 completed questionnaires about fatal risks.

In this questionnaire, risks were expressed as X in 100,000, and depicted using a grid of squares. People were first shown, as an example, the risks of dying for various causes (all causes, heart disease, stomach or esophageal cancer, traffic accident) for a 50-year-old. They were then asked to assess subjectively their risk of dying for any cause, and in a road-related traffic accident. They were also asked directly to report their WTP for a reduction in each of these two risks. Unlike the previous studies, the payment question used an open-ended format, resulting in observations about WTP on a continuous scale.

In this paper, we focus on the WTP for a reduction in the risk of dying in a traffic accident. This risk reduction is expressed as a proportion (10%, 30%, or 50%, depending on the questionnaire version⁴) of the baseline risk. The risk reduction is a private commodity (safety equipment and preventive health care) and is valid for one year. A reminder of the respondent's budget constraint is provided.

⁴ Respondents were randomly assigned to one of these possible risk reductions.

Table 1. Mortality Risk Studies examined in this paper. All monetary figures in US dollars, unless otherwise indicated.

Study	Description and VSL
Johannesson et al. (1997)	<p>Telephone survey of Swedes aged 18-74. Dichotomous-choice questions about WTP for 2 in 10,000 reduction in their risk of dying (from all causes).</p> <p>VSL \cong \$4.5 million.</p>
Persson et al. (2001)	<p>Mail survey in Sweden. Elicits WTP for X% reduction in the risk of dying in a road-traffic accident. Subjective baseline risks. Open-ended WTP questions.</p> <p>VSL = \$2.84 million (based on WTP for 2 in 100,000 risk reduction).</p>
Krupnick et al. (2002)	<p>Survey of persons aged 40-75 years in Hamilton, Ontario. Self-administered computer questionnaire, centralized facility. Dichotomous-choice payment questions with dichotomous-choice follow-up question.</p> <p>VSL = Can \$1.2 to 2.8 million (US \$ 0.96 to 2.24 million).</p>
Alberini et al. (2004)	<p>US national survey conducted over Web-TV. Dichotomous-choice payment questions with dichotomous-choice follow-up question.</p> <p>VSL = \$700,000 to \$1.54 million (based on 5 in 1000 risk reduction)</p>

III. Estimation of VSL with Dichotomous Choice Data

Dichotomous-choice payment questions ask respondents whether they would be willing to pay a specified amount of money to obtain the risk reduction stated to them in the questionnaire. The amount of money (usually termed “the bid”) is randomly assigned to the respondent out of a list of preselected values, and is varied across respondents. Respondents are offered two possible response categories: “yes” and “no.”⁵ Their

⁵ When the risk reduction is delivered by a public program, the payment question is often phrased in terms of vote in a referendum on a ballot. The respondent is told that the program would be implemented only if there are a majority of votes in favor of the program, and that the cost of the program—usually, in the form of an income tax—for his household is \$X. If a majority is not reached, the program is abandoned, and no

responses imply that WTP is greater (“yes”) or less than (“no”) the bid, but the exact WTP amount is not observed.

Estimates of mean WTP and other welfare statistics (e.g., median WTP) can be obtained by estimating binary data models that rely on this mapping from the unobserved WTP amount to the response to the payment question. For example, if latent WTP is normal (logistic) with mean μ and scale σ , a probit (logit) model is estimated where the dependent variable is a dummy indicator that takes on a value of one if the response to the payment question is a “yes” and zero otherwise, and the right-hand side includes the intercept and the bid. Cameron and James (1987) show that mean/median WTP is equal to $-\alpha/\beta$, where α and β are the probit (logit) intercept and slope, respectively.⁶

In the remainder of this section, we examine the sensitivity of the estimates of WTP (and hence VSL) based on dichotomous choice data to the distribution WTP is assumed to follow, the welfare statistic one wishes to work with, and the procedure used by the researcher in computing it. To illustrate the consequences of assumptions and procedures, we use the data in Johannesson et al. (1997), a telephone survey of Swedes aged 18-74 about their WTP for a 2-in-10,000 reduction in their risk of dying over the next year.

Respondents were informed about the chance of dying for a person of their age and gender over the next year, and were queried about their WTP to reduce that risk

additional income taxes are incurred by households. The two possible response categories are “in favor” and “against” the program.

⁶ Symmetric distributions like the normal and logistic imply that mean WTP is equal to median WTP. If WTP is assumed to be a lognormal, the probit equation is amended by replacing the bid in the right-hand side of the model with its logarithmic transformation. Median WTP is equal to $\exp(-\alpha/\beta)$, and mean WTP is equal to $\exp(0.5 \cdot (1/\beta)^2 - \alpha/\beta)$. With a Weibull distribution, a binary choice model is estimated where $\Pr(\text{yes}|\text{B}) = \exp(-(B/\sigma)^\theta)$, where θ and σ are the scale and shape parameters, respectively, of the Weibull variate. Mean WTP is $\sigma \cdot \Gamma(1/\theta + 1)$ and median WTP is $\sigma \cdot (-\ln(0.5))^{1/\theta}$, where $\Gamma(\cdot)$ is the gamma function.

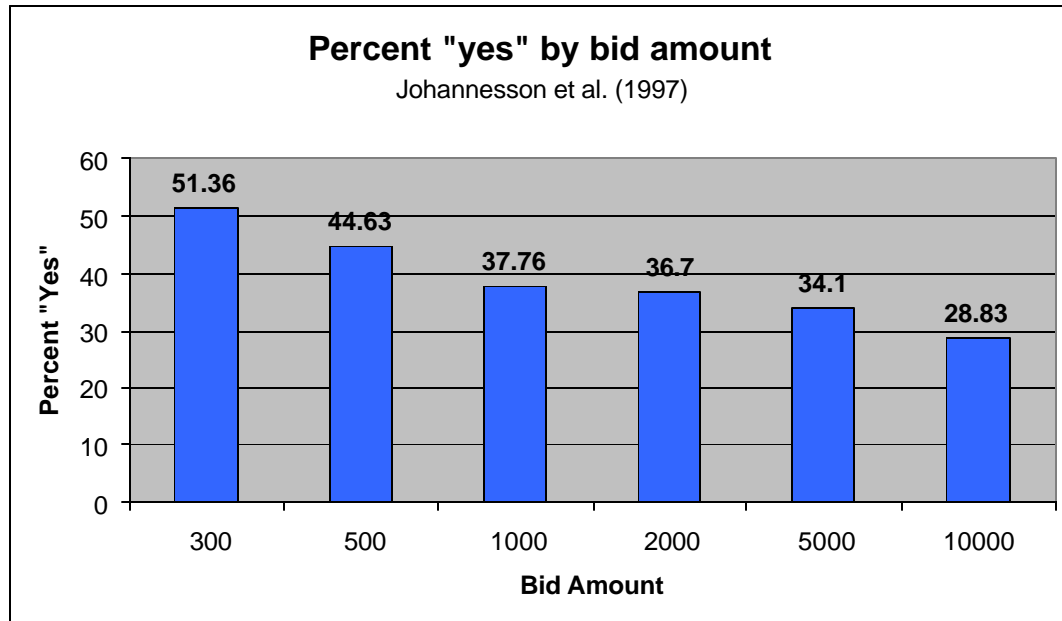
using dichotomous choice questions. The bid values ranged between 300 and 10,000 SEK (about \$40 to \$1400, implying VSL values of \$200,000 to \$7 million). Johannesson et al. estimate mean WTP to be 6300 SEK, or about \$900.

A. WTP Responses and WTP Distribution

We begin our examination of the data from the Johannesson et al. study by checking whether (i) the percentage of “yes” responses decline with the bid amount, and (ii) the bids cover a reasonably wide portion of the range of WTP values.⁷ As shown in Figure 1, the percentage of “yes” responses declines from 51.36% at the lowest bid amount, 300 SEK, to 28.83% at the higher bid amount 10,000 SEK, satisfying the first of these two requirements. Figure 1 also implies that all bids are greater than median WTP, failing to satisfy requirement (ii), and raising concerns about the stability of the estimates of WTP. Median WTP is pegged between 300 and 500 SEK.

⁷ Cooper (1993) emphasizes the importance of using a vector of bids that covers the entire range of possible WTP values. Kanninen (1993) and Alberini (1995) derive c-optimal and d-optimal designs for dichotomous-choice CV surveys that rely on only two bid values.

Figure 1.



We estimate mean and median WTP under four alternative distributional assumptions. Mean and median WTP are derived directly from the estimated parameters of the binary-response models, as explained above. Results are reported in table 2.

The most surprising result of table 2 is that the estimates of mean and median WTP are *negative* when WTP is assumed to follow the normal or the logistic distribution. The model based on the normal distribution predicts that 54% of the respondents have negative WTP values. Using the Weibull and lognormal distributions, which admit only non-negative values of WTP and fit the data better,⁸ circumvents this problem, but results in a large discrepancy between median and mean WTP. Mean WTP is very large. The median WTP amounts predicted by the two distributions are relatively close to one another (239 and 250 SEK for Weibull and lognormal, respectively), but both are less

⁸ The lognormal and Weibull distribution results in higher Akaike Information Criterion (AIC) values. The AIC is computed as the log likelihood minus the number of parameters to be estimated, and is frequently used in applied work to assess the fit of a model.

than what would be inferred by examining the responses to the payment questions, and less than the smallest bid value offered to the respondents in the study.

Table 2. Mean and Median WTP for various distributional assumptions (Johannesson et al. study, 1997).

	Normal	Logistic	Weibull	Lognormal
Mean WTP	-2096.08	-2007.75	2,894,292	Infinity
Median WTP	-2096.08	-2007.75	238.39	254.30
Log L	-1349.19	-1349.10	-1344.01	-1343.84

Clearly, these figures are very different from those reported by Johannesson et al. (1997), who rely on a completely different procedure for estimating mean WTP. Specifically, they start with fitting a logit model, which implicitly admits negative WTP values, but compute mean WTP as the area under the survival curve for *positive* WTP values:

$$(1) \quad \int_0^{\infty} [1 - G(y)] dy ,$$

where $G(y)$ is the cdf of WTP. When WTP is a logistic variate with mean μ and scale σ , it can be shown that (1) is equal to $(-1/\beta) \ln[1 + \exp(\alpha)]$, where $\alpha = \mu/\sigma$ is the intercept and $\beta = -1/\sigma$ is the slope of the logit model. Because negative WTP are implicitly allowed, but subsequently ignored in the procedure for computing mean WTP, which is akin to calculating the mean of a tobit variate even though no tobit model was estimated in the first place, (1) is expected to produce a higher estimate of mean WTP than $-\alpha/\beta$.

B. Changing the Procedure for Estimating Mean WTP

In table 3, we experiment with alternative calculations of mean WTP that are variants on four basic procedures. The first procedure follows Cameron and James (1987). We fit a probit or logit model of the “yes” or “no” responses to the payment questions, and compute mean WTP as

$$(2) \quad m_1 = -\alpha / \beta.$$

The second is the procedure followed by Johannesson et al., who fit a logit model but effectively disregard the portion of the distribution corresponding to negative values. If WTP follows the logistic distribution, this yields:

$$(3) \quad m_2 = (-1/b) \ln[1 + \exp(a)].$$

Our third procedure continues to rely on the fact that mean WTP is the area under the survival curve, i.e., $[1 - F(a + by)]$, where $F(\cdot)$ is the cdf of the standardized WTP variate. In earlier applications of the CV method, researchers estimated mean WTP by computing the area under the fitted survival curve up to the largest bid amount offered in the survey (10,000 SEK in the Johannesson et al. study). Our third estimate of mean WTP is thus:

$$(4) \quad m_3 = \int_0^{B_{\max}} [1 - F(a + by)] dy.$$

Finally, Chen and Randall (1998) and Creel and Loomis (1997) describe semiparametric approaches to modeling the WTP responses. Specifically, they propose to estimate m_3 by improving the fit of $F(\cdot)$ through augmenting its argument to include terms

such as the sine and cosine transformations of the bid and of other regressors, in the spirit of fast Fourier transform approximations.⁹ The argument of $F(\bullet)$, therefore, becomes:

$$(5) \quad z = \mathbf{x}\mathbf{b} + \sum_{a=1}^A \sum_{j=1}^J [u_{ja} \cos(j\mathbf{k}_a s(\mathbf{x})) - v_{ja} \sin(j\mathbf{k}_a s(\mathbf{x}))],$$

where \mathbf{x} is a vector that includes the bid and other determinants of WTP. For a subset, or all, of these variables (the dimension of this subset being A), we introduce a scaling function $s(\mathbf{x})$. This scaling function subtracts the minimum value of \mathbf{x} , divides the result by the maximum value of \mathbf{x} (thus forcing the rescaled variables to be between zero and 1), and then multiplies it by $(2\pi - 0.00001)$. For this rescaling function to be possible, there must be at least three distinct values for \mathbf{x} , which rules out applying this transformation to dummy variables. The \mathbf{k} s are vectors of indices, and the u s are parameters to be estimated. Chen and Randall (1998) and Creel and Loomis (1997) suggest that for most dichotomous choice CV survey applications it is sufficient to consider $J=1$, in which case z is simplified to:

$$(6) \quad z = \mathbf{x}\mathbf{b} + \sum_{a=1}^A [u_a \cos(s(\mathbf{x})) - v_a \sin(s(\mathbf{x}))].$$

We apply the semiparametric approach defined by equations (5) and (6) to the Johannesson et al. data, where $F(\bullet)$ is the standard normal (logistic) cdf, and z , the argument of the standard normal (logistic) cdf, includes an intercept, the bid and its sine and cosine transformations (after rescaling). Formally, we compute mean WTP as

$$(7) \quad \int_0^{\bar{B}} [1 - F(\mathbf{a} + \mathbf{b}y + \mathbf{d} \sin(y') + \mathbf{g} \cos(y'))] dy,$$

⁹ Chen and Randall also consider polynomial terms in the variables \mathbf{x} .

where $y' = (2\mathbf{p} - 0.00001) \cdot \left[\frac{y - B_{\min}}{B_{\max}} \right]$, B_{\min} and B_{\max} are the smallest and largest bid

amounts used in the survey, and \bar{B} is the upper limit of the integration.

Following Creel and Loomis (1997), we first set \bar{B} equal to the largest bid amount used in the study ($\bar{B} = B_{\max}$). This defines our estimate m_4 of mean WTP. We subsequently compute m_5 by letting \bar{B} in (6) tend to infinity.¹⁰

The results from these alternative calculations are shown in table 3. Table 3 shows that the largest change in estimated mean WTP occurs when going from m_1 —which yields a negative mean WTP—to approaches m_2 - m_5 , which restrict integration to the positive semiaxis (or a portion of it). Using the standard normal or the standard logistic cdf gives similar results (table 3, second and third rows). The estimate of mean WTP is sensitive to the upper limit of integration. As shown in table 3, third and fifth rows, for the probit model, when the upper limit of integration is B_{\max} , mean WTP is roughly half the figure that is obtained by letting \bar{B} tend to infinity. A similar comparison for the logit model (table 3, second and fourth rows) confirms these findings.

The semiparametric approach results in an estimated mean WTP similar to that of regular probit and logit models when $\bar{B} = B_{\max}$, but that is considerably more conservative than the regular probit and logit equations when \bar{B} tends to infinity. Assuming that WTP is normally distributed and that a conventional probit model is fit,

¹⁰ Cooper (2002) points out that it is not clear a priori which of these two estimates— m_4 or m_5 —is greater. This is because, unless additional restrictions are imposed, when we adopt the semiparametric approach $F(z)$ as defined in (7) can no longer be interpreted as the cdf of the standardized WTP variate. Also see Crooker and Herriges (2004) for a comparison between the Chen and Randall approach and other models of dichotomous choice CV responses.

m_2 is 6434 SEK, whereas m_5 based on the semiparametric probit is equal to 4338 SEK (a 33% reduction).

A comparison between the probabilities of “yes” responses predicted by the semiparametric and conventional probit at various bid values suggests that the former outperforms the latter. For example, the former predicts that when the bid is 300 SEK the probability of “yes” is 0.5172, which is closer to the relative frequency (0.5136) than the prediction from the conventional probit (0.4794). When the bid amount is 10000 SEK, the semiparametric probit predicts that the probability of a “yes” is 0.2857 against 0.2669 from the conventional probit. For comparison, the empirical frequency is 0.2883.

Table 3. Alternative procedures for computing mean WTP.
Johannesson et al. (1997) data.

<u>Approach</u>	<u>Distribution F()</u>	<u>Mean WTP (in SEK)</u>
m_1 (Cameron and James, 1987)	Standard logistic	-2007
m_2 (Johannesson et al., closed-form expression)	Standard logistic	6849
m_2 (Numerical integration of the survival function to infinity)	Standard normal	6434
m_3 (Numerical integration of the survival function up to max. bid)	Standard logistic	3522
m_3 (Numerical integration of the survival function up to max. bid)	Standard normal	3528
m_4 (Creel and Loomis (1998) semiparametric approach. Numerical integration up to max. bid)	Standard normal; probit model with bid, sin(bid) and cos(bid)	3732
m_5 (Creel and Loomis (1998) semiparametric approach. Numerical integration with $B_{\max} = \infty$)	Standard normal; probit model with bid, sin(bid) and cos(bid)	4339

C. The Effect of Regressors

Would we obtain similar results in situations where regressors are included in the model, and mean WTP is calculated conditionally on specific values of the regressors? This question is appropriate, for example, when seeking to answer the question of how the WTP for a risk reduction varies with age. Epidemiological evidence (e.g., Pope et al., 1995) suggests that the majority of the lives saved by air quality regulations and environmental policies are those of the elderly, and some observers have argued that older people should be willing to pay less for a risk reduction—and their VSL should be lower—mirroring their fewer remaining life years. Economic theory, however, does not offer unambiguous predictions about the effect of age on WTP (Alberini et al., 2004).¹¹

Johannesson et al. run a logit regression that includes age and age squared, plus gender and education dummies, income and the respondent's quality-of-life rating,¹² and report finding a quadratic relationship between age and WTP that peaks when the individual is about 40 years old. To check the sensitivity of these results to the procedure used in the calculation, we ran logit and probit models with their same regressors (or subsets of them), and predicted mean WTP at different ages using approaches m_2 and m_4 . In both cases, we let \bar{B} tend to infinity.

The results of these calculations are shown in Figure 2, panels (a)-(d). Panel (a) plots m_2 against age, confirming Johannesson et al.'s finding: the relationship between

¹¹ Because a large proportion of the lives saved appear to be those of the elderly, there has been much recent debate about whether the VSL should be lower for the elderly to reflect their fewer remaining life years. In the US, the Office of Management and Budget recently repudiated making such adjustment for age, on the grounds of insufficient evidence that the VSL is lower for elderly persons (Skrzycki, 2003).

¹² In the Johannesson et al. survey, respondents were asked to rate their quality of life on a scale from 1 to 10, where 1 represents the worst possible quality, and 10 is the best possible quality.

age and WTP is an inverted U that peaks at age 40.¹³ Similar results are observed when, as shown in panel (b), the logit model is replaced by a probit, and mean WTP is computed as $\int_0^{\infty} [1 - \Phi(x_i \mathbf{a} + \mathbf{b}y)] dy$, where $\mathbf{x}_i' = [1 \quad \text{age} \quad \text{age}^2]$, α is the vector of probit coefficients on these variables, and $\beta = -1/\sigma$.

In Figure 2, panel (c), we compare the predictions based on the conventional logit model with those from semiparametric probit models. Probit Fourier 1 is based on a probit regression where the right-hand side variables are the bid, age and age squared, and trigonometric functions of these variables. In this case, the shape of the relationship between WTP and age is no longer an inverted-U. Moreover, this approach produces estimates of mean WTP that are consistently smaller than those from the conventional probit model. The curve labeled Probit Fourier 2 is based on a similar model, except that the sine and cosine transformations are applied only to the bid. This time, the relationship between WTP and age resumes its quadratic shape, but the estimated WTP values remain consistently lower than those of the regular probit model.

Figure 2, panel (d) displays the results based on probit models that are similar to those used for panel (c), except that more regressors—household income, the quality-of-life rating reported by the respondent, a gender dummy and an educational attainment dummy—are entered in the right-hand side of the model. The curves labeled Probit Fourier 1 and Probit Fourier 2 are different from one another in that the former includes trigonometric functions of all of the continuous variables, while the latter includes only the sine and cosine transformations of the bid.

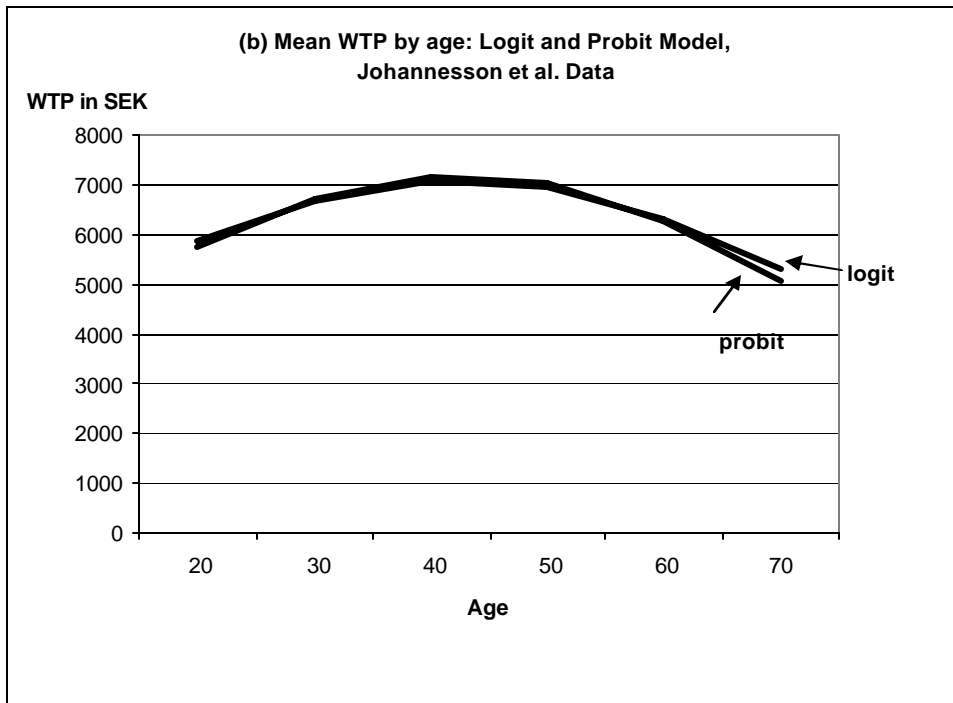
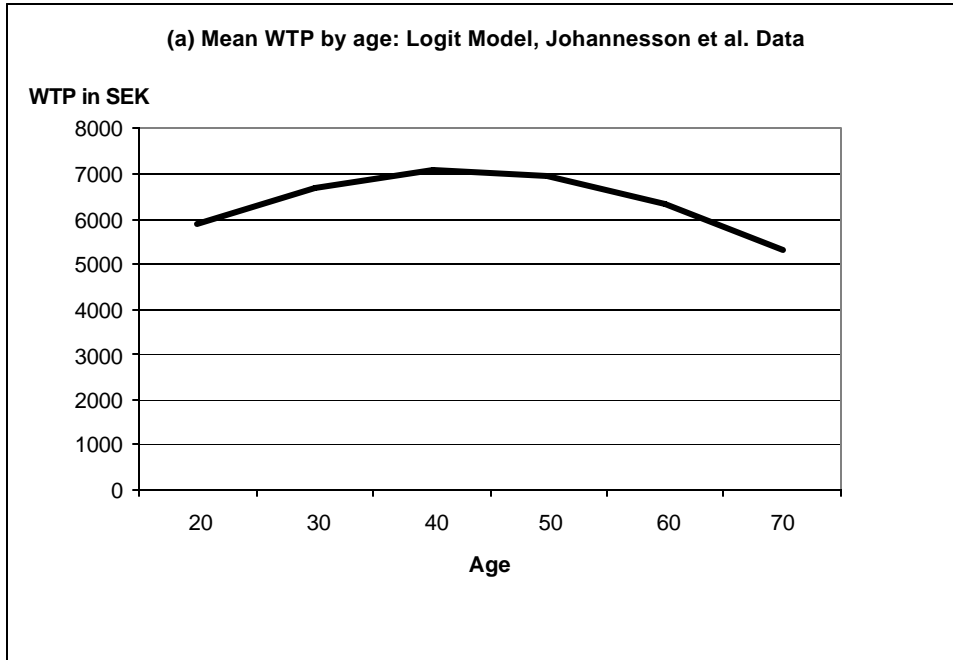
¹³ Dummy indicators for the “yes” or “no” responses to the payment questions were regressed on an intercept, the bid amount, age and age squared.

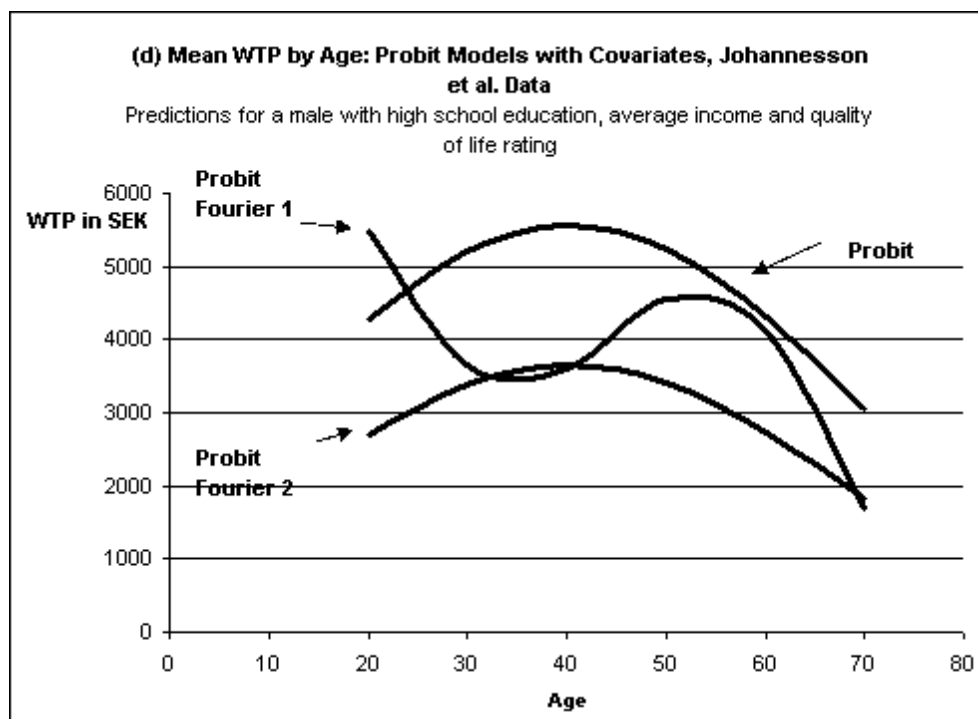
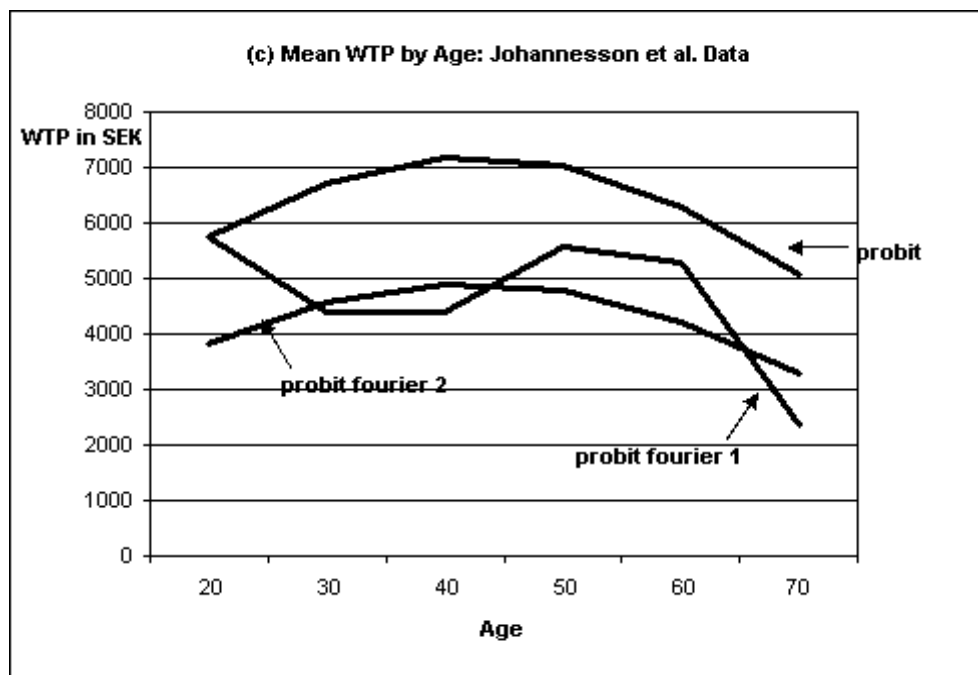
The mean WTP figures plotted in panel (d) refer to a 50-year-old male (SEX=1) with high school education (DEDU=1), the average household income of the sample (24,490 SEK), and the same quality-of-life rating as the average respondent (7.34 on a scale from 1 to 10). As in panel (c), Probit Fourier 1 results in a non-monotonic relationship between WTP and age, while Probit Fourier 2 implies a quadratic relationship. Both predict lower WTP figures than the conventional probit model.

Figure 2.

Legend: in panel (c), Probit Fourier 1 includes bid, age, age squared, and sine and cosine functions of these variables; Probit Fourier 2 includes bid, $\sin(\text{bid})$, $\cos(\text{bid})$, age and age squared.

In panel (d), Probit Fourier 1 includes bid, age, age squared, income, quality of life rating, a gender dummy and an education dummy, plus sine and cosine terms of all continuous variables. Probit Fourier 2 includes bid, $\sin(\text{bid})$, $\cos(\text{bid})$, age, age squared, income, quality of life rating, a gender dummy and an education dummy.





In sum, panels (c) and (d) in Figure 2 suggest that there are no easily discernible patterns, and that claims about the relationship between age and WTP, and the magnitude of WTP at various ages, are not robust and may be an artifact of restrictive assumptions.

We further investigate this matter by switching to a lognormal distribution for WTP and to median WTP for specific ages, which we expect to result in more conservative estimates. The results, shown in table 4, suggest that the lognormal model implies a quadratic, inverted-U relationship between age and WTP. It also suggests, however, that the curvature of the relationship is *much sharper* than that predicted by Johannesson et al. For example, the WTP of a 70-year-old for a reduction in risk of 2 in 10,000 is only 90 SEK, or only about 20% of the WTP predicted for a 40-year-old person (440 SEK).

Taken together with the evidence from the semiparametric approach, these results suggest that detecting the shape of the relationship between WTP and a regressor of interest depends crucially on, and is very sensitive to, three factors: (i) the distribution assumed for WTP, (ii) the welfare statistic used (mean or median WTP), and (iii) the procedure used for computing it.

Table 4. The relationship between age and WTP for a risk reduction:
lognormal WTP and median WTP v. Johannesson et al. logit and truncated mean WTP.

Age	Johannesson et al, 1997.		Alternative calculation using log normal WTP.	
	Mean WTP in SEK	Implied VSL in million SEK	Median WTP in SEK	Implied VSL in million SEK
20	6100	30.3	137.18	0.672
30	6900	34.6	307.31	1.505
40	7200	36.1	440.77	2.160
50	6900	34.3	404.75	1.983
60	6000	29.8	237.97	1.166
70	4600	23.3	89.57	0.439

IV. Treatment of Outliers

In this section, we investigate the effect of outliers on the estimates of WTP. We begin by tackling the problem of identifying outliers in dichotomous-choice CV surveys, and explore the effect of including or excluding these observations from the sample. We use logit regressions of the WTP responses on individual characteristics to classify observations as outliers. Because outliers can be caused, among other reasons, by a number of undesirable response effects, we then examine whether it is possible to estimate “structural” models of these response effects.

A. Outliers

Collett (1991) defines as outliers as “observations that are surprisingly far away from the remaining observations in the sample,” and points out that such values may occur as a result of measurement errors, execution error (i.e., use of a faulty experimental procedure), or be a legitimate, if extreme, manifestation of natural variability.

Our first order of business is to define outliers when the variable of interest is binary, as is the case with the responses to dichotomous-choice CV questions. Copas (1988) defines an outlier as an observation for which we predict a low probability of a one (zero), but we do observe a one (zero).

We use the Johannesson et al. (1997) data to check (i) how many observations could be classified as outliers according to several alternative cutoff levels, and (ii) by how much mean WTP would change if these outliers were excluded from the sample. Specifically, we wish to see for how many observations the predicted probability of a “yes” is less than 0.05, 0.10, etc., but the response to the payment question is a “yes.” The predicted probability is based on Johannesson et al.’s logit regression of the “yes” or “no” response indicator on respondent age, age squared, income, an education dummy, and a quality-of-life rating subjectively reported by the respondent in the interview:

$$\hat{p}_i = [1 + \exp(-(x_i \hat{\mathbf{a}} + \hat{\mathbf{b}} \cdot B_i))]^{-1},$$

where $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ are the estimated logit coefficients, \mathbf{x} is a vector of regressors, and B is the bid assigned to respondent i .

For ease of comparison, we use the same procedure for estimating mean WTP as in Johannesson et al.’s work (see section III). The resulting mean WTP figures are reported in column (C) of table 5. We also fit a binary data model of the responses based on an alternate distribution—the Weibull—and report estimates of mean and median WTP based on the latter in columns (D) and (E) of table 5, respectively.

Table 5. Outliers in the Johannesson et al. data (based on logit regression, n=1660):
All WTP figures in SEK.

(A) Definition of outlier	(B) How many?	(C) Johannesson et al. procedure Mean WTP*	(D) Weibull: Mean WTP*	(E) Weibull: Median WTP*
No outliers identified	None	6732	2.894 million	238
Prob(yes) ≤ 0.05 and yes observed	None	6732	2.894 million	238
Prob(yes) ≤ 0.10 and yes observed	None	6732	2.894 million	238
Prob(yes) ≤ 0.20 and yes observed	5	6141	1.150 million	302
Prob(yes) ≤ 0.25 and yes observed	26	4846	193,481	338
Prob(yes) ≤ 0.30 and yes observed	59	3767	36,114	369

*: Welfare statistics after excluding outliers.

Table 5 shows that outliers according to the Copas' definition were found only when the cutoff for identifying an outlier was set to 0.15 or higher. When the cutoff is set to 0.25, for example, a total of 26 observations would be considered outliers, and dropping them from the usable sample would reduce Johannesson et al.'s mean WTP from 6732 to 4846 SEK—a 30% reduction. A cutoff of 0.30 results in the exclusion from the usable sample of 33 more individuals, and in a further reduction of mean WTP to 3767 SEK—a 45% reduction.

As shown in column (D) of table 5, using a Weibull distribution generally results in implausibly large mean WTP values. The mean WTP, however, does get smaller when outliers are excluded from the sample. By contrast, median WTP (shown in column (E)), which is 338 SEK for the full sample, increases slightly when outliers are omitted from

the sample, which suggests that the skewness of the distribution of WTP has become a little less pronounced.

B. Undesirable Response Effects: Yea-saying, Nay-saying, and Random Responses

Contingent valuation studies about mortality risk reduction rely crucially on the respondent's comprehension of the risk and risk reductions being valued. Many recent survey questionnaires deploy visual aids and practice questions about risks, but, despite these efforts, it is possible that some respondents still remain confused about the commodity being valued, and that their answers to the payment questions may be affected by undesirable response effects.

Carson (2000) describes three types of undesirable response effects that may occur in dichotomous-choice CV surveys. The first is yea-saying, whereby a respondent answers "yes" to the bid question with probability 1, regardless of the bid amount. This may be done in an effort to please the interviewer, or in the hope of terminating the interview sooner.

The opposite phenomenon is nay-saying, which occurs when the respondent answers "no" to the payment question with probability 1, regardless of the bid amount. Respondents engaging in nay-saying may dislike new public programs and new taxes, or might be afraid of committing to something they do not fully understand.

It is also possible that some people give completely random responses, answering "yes" to the payment question with probability 0.5 (and hence, "no" with probability 0.5), regardless of the bid amount. Completely random responses may be due to confusion

about the scenario, failure to understand the commodity being valued, no interest in the survey, poorly written survey questions or survey materials, or simply a data entry error.

In practice, not all respondents in a contingent valuation survey will be subject to these undesirable effects. To accommodate for this possibility, we consider discrete mixtures. For simplicity, attention is restricted to discrete mixtures with two components, where a fraction of the sample ($\alpha \cdot 100\%$) is affected by one of these undesirable response effects, while the remainder answers the payment questions in the usual fashion (i.e., by saying “yes” if latent WTP is greater than the bid, and “no” otherwise). The researcher’s problem is that—unless respondent or interviewer debriefs are used—it is not possible to tell from which component of these two populations the respondent is drawn. This requires estimating a (discrete) mixture of distributions.

Figures 3 and 4 show how mixtures alter the estimated survival curve of WTP. Figure 3 depicts the true and observed survival curve when one of the two mixing components is yea-saying. (By true survival curve, we mean the survival curve that refers to the non-degenerate component of the mixture.) Yea-saying raises the observed survival curve at every bid value, which implies that mean WTP will be overestimated. The extent of this positive bias depends on severity of the yea-saying (i.e., on the value of α).

The converse will be true with a nay-saying component. Figure 4 shows that if the sample is contaminated with a small number of people who answer the payment question in a completely random fashion, the observed survival curve is tilted, crossing the true survival function from below at the median, and remaining above it for higher bid values. This implies that both the mean and variance of WTP will be overestimated, but that the

median WTP should not be affected by the presence of completely random responses, as long as the mixture contains only these two components.

C. Likelihood Functions for Two-component Mixtures

Absent any other undesirable response effects, the log likelihood function with dichotomous choice responses is:

$$\begin{aligned}
 (8) \quad \log L &= \sum y_i \log \Pr(y_i = 1 | B_i) + (1 - y_i) \log \Pr(y_i = 0 | B_i) \\
 &= \sum y_i \log [1 - F(B_i; \mathbf{q})] + (1 - y_i) \log F(B_i; \mathbf{q})
 \end{aligned}$$

where y is an indicator that takes on a value of 1 if the response is a “yes” and zero if it is a “no,” B denotes the bid, and $F(B_i; \mathbf{q})$ is the cdf of WTP evaluated at the bid value, θ being the vector of parameters indexing F .

In this paper, we consider a two-component mixture where $(1-\alpha) \cdot 100$ percent of the population answers the payment questions according to the usual assumption (“yes” if WTP is greater than the bid, “no” otherwise). The distribution of WTP for this component of the mixture is non-degenerate. If the remaining $\alpha \cdot 100$ percent of the population consists of yea-sayers, who answer “yes” to the payment question with probability 1,¹⁴ the probability of observing a “yes” in equation (8), $\Pr(y_i = 1 | B_i)$, is equal to:

$$(9) \quad \Pr(y_i = 1 | B_i) = (1 - \alpha) \cdot [1 - F(B_i; \mathbf{q})] + \alpha \cdot 1.$$

¹⁴ In practice, one would expect the data from a contingent valuation survey to come from mixtures with more than two components. The simple cases considered in this paper should, therefore, be interpreted as the situations where the researcher stands his or her best chance to identify the components.

“No” responses must come from the non-degenerate component of the mixture:

$$(10) \quad \Pr(y_i = 0 \mid B_i) = (1 - \mathbf{a}) \cdot F(B_i; \mathbf{q}) .$$

If the sample is described by a discrete mixture with nay-sayers, the probability of a “yes” is:

$$(11) \quad \Pr(y_i = 1 \mid B_i) = (1 - \mathbf{a}) \cdot [1 - F(B_i; \mathbf{q})] ,$$

and the probability of observing a “no” is:

$$(12) \quad \Pr(y_i = 0 \mid B_i) = (1 - \mathbf{a}) \cdot F(B_i; \mathbf{q}) + \mathbf{a} \cdot 1 .$$

Finally, if the sample is a mixture comprising persons who answer in a completely random fashion, the probabilities of a “yes” and “no” response are

$$(13) \quad \Pr(y_i = 1 \mid B_i) = (1 - \mathbf{a}) \cdot [1 - F(B_i; \mathbf{q})] + \mathbf{a} \cdot 0.5 , \text{ and}$$

$$(14) \quad \Pr(y_i = 0 \mid B_i) = (1 - \mathbf{a}) \cdot F(B_i; \mathbf{q}) + \mathbf{a} \cdot 0.5 ,$$

respectively.

Figure 3. Effect of yea-saying.

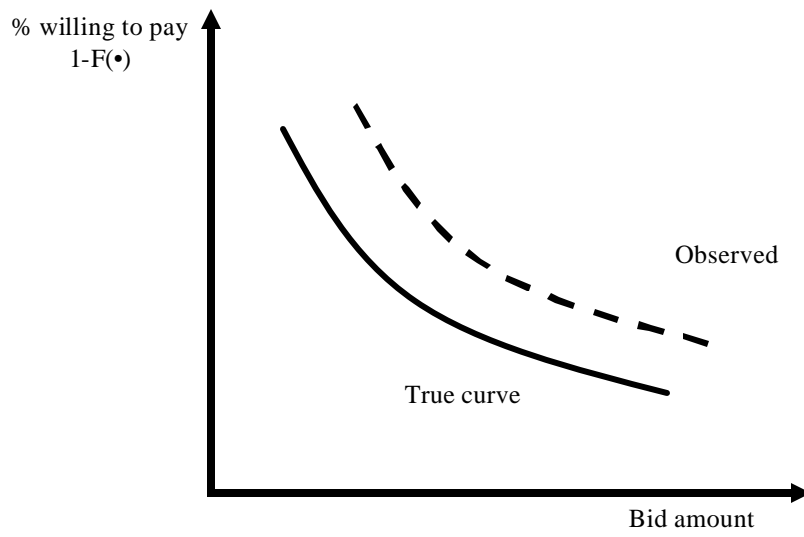
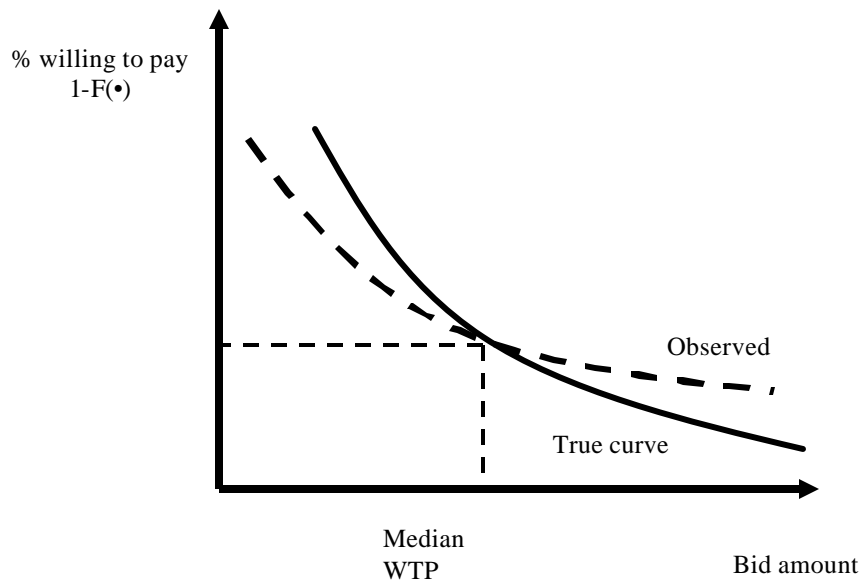


Figure 4. Effect of Completely Random Responses.



C. Application

We apply these two-component mixtures to the data from the CV survey conducted in Hamilton, Ontario, by Krupnick et al. (2002). Participants in this study took a self-administered computer questionnaire at a centralized facility. After a probability tutorial, the study participants were shown the risk of dying for all causes for a person of their age and gender.

They were asked to value a total of three risk reductions by answering payment questions in a dichotomous-choice format with follow-up.¹⁵ Each of these risk reductions would be taking place over 10 years, and would be delivered by a hypothetical product. The risk reduction was context-free, in that it was not associated with air pollution reductions or another public program.

In this paper we focus on WTP for the 5 in 1000 risk reduction. We assume that WTP is a Weibull variate, and begin with fitting the models described by equations (9)-(14) to the responses to the initial payment questions. In a subsequent round of estimation, we amend equations (9)-(14) to accommodate for the responses to follow-up questions (double-bounded models). Table 6 reports mean and median WTP for the component of the mixture that behaves following the usual economic paradigm (after filtering out the degenerate components).¹⁶

As shown in table 6, panel (i), in our mixture based on single-bounded models we find no evidence of yea-saying: in column (ii), the estimated α is identically equal to zero. By contrast, column (iii) suggests that as much as 26% of the sample may be

¹⁵ The three risk reductions were 5 in 1000 and 1 in 1000 occurring over the next 10 years, and 5 in 1000 beginning at age 70.

¹⁶ We use the constrained maximum likelihood routine in GAUSS to force α to lie between 0 and 1.

comprised of nay-sayers.¹⁷ Column (ii) displays an even more implausible result: in the mixture with completely random responses, we estimate that over 50% of the sample answers the payment questions in a completely random fashion.

Because we do not know whether a specific individual does or does not answer the payment question at random, this implies that each observation has a probability equal to 0.51 of being a completely random response. This finding is in sharp contrast with the answers to debriefing questions and the good internal consistency and validity shown by study participants, raising doubts about the ability of single-bounded models of WTP to capture the correct proportion of subjects engaging in degenerate response mechanisms.

In theory, double-bounded estimation should produce more reliable results, thanks to the narrower intervals around the respondent's unobserved WTP amounts afforded by the follow-up payment question. Table 6, panel (ii) reports estimation results based on double-bounded models. While any evidence of nay-saying seems to have vanished (column (iii)), almost two-thirds of the respondents are predicted to be answering the payment questions in a completely random fashion (column (ii)). Clearly, this result is not plausible.

To seek for an explanation for this finding, we turn to evidence from Monte Carlo simulations. Alberini and Carson (2001) report that mixture models like (9)-(14) give reliable results only when (a) $F(\bullet)$ is correctly specified, (b) the mixture type is correctly specified, and (c) α is not too small. In most other cases, the estimated coefficient $\hat{\alpha}$ appears to pick up any divergence between the assumed $F(\bullet)$ and the true distribution, or

¹⁷ We view this as a somewhat alarming result, but this percentage is comparable to the probability of yeasaying (20%) estimated by Kanninen (1995) using double-bounded data from a contingent valuation survey about wildlife and wetland habitat protection in California's San Joaquin Valley.

any other misspecification in the type of mixture. We conclude that it may be difficult to estimate mixtures using existing CV data about mortality risk reductions, and that a study capable of identifying the components of a mixture would probably have to be designed specifically for that purpose.

Table 6.
Discrete Mixture Models:
Results for Krupnick et al. (2002) (Canada study) Bootstrap standard errors in parentheses.

I. using only the responses to the initial payment questions (single-bounded).

	(i) No mixture	(ii) Mixture with completely random responses	(iii) Mixture with nay- sayers	(iv) Mixture with yea-sayers
α (probability of yea-sayers, nay- sayers, etc.)	--	0.5108 (0.13)	0.2644 (0.05)	--
Mean WTP (\$)	1176.94 (306.55)	594.55 (n/a)	969.57 (n/a)	1176.94 (n/a)
Median WTP (\$)	445.99 (44.02)	551.04 (146.05)	859.61 (134.10)	445.99 (44.02)

II. using the responses to the initial and follow-up payment questions (double-bounded).

	(i) No mixture	(ii) Mixture with completely random responses	(iii) Mixture with nay- sayers	(iv) Mixture with yea-sayers
α (probability of a yea-sayer, nay- sayer, etc.)	--	0.6577 (0.0448)	--	--
Mean WTP (\$)	826.41 (70.85)	831.24 (70.92)	826.41 (70.85)	826.41 (70.85)
Median WTP (\$)	323.83 (20.96)	323.75 (21.10)	323.83 (20.96)	323.83 (20.96)

V. Internal Validity: Scope Tests and Endogenous Risks.

A. Subjective Risks and Risk Reductions

In this section, we consider the situation where respondents are asked to assess their subjective baseline risks (Gerking et al., 1988; Johannesson et al., 1991;¹⁸ Persson et al., 2001). If the risk reduction is expressed as a specified proportion of the baseline risk (e.g., 20%), then the risk reduction varies across respondents, allowing one to test for sensitivity of WTP with respect to the size of the risk reduction—the so-called “scope” effect.¹⁹

In this section, we ask three related questions. First, should WTP and risk reduction be treated as endogenous in such studies? Second, does this affect conclusions about the “scope” effect? Third, does treating risk reduction as endogenous with WTP affect our conclusions about whether subjects respond to absolute or relative risk reductions when they announce their WTP amounts (Baron, 1997; McDaniel, 1992)?

To answer these questions, we use the data from the Persson et al. (2001) study about transportation safety. Persson et al. conducted a mail survey of 18-74-year-old Swedes in 1998. Respondents were asked to estimate their own risk of dying from any cause (EIGRISK), and to value a specified reduction in this risk. The reduction was

¹⁸ Johannesson et al. (1991) ask hypertensive patients in a health care facility to assess both baseline risks and risk reductions associated with hypertension medication. In this case, it seems reasonable to expect that WTP and risk reductions should be endogenous.

¹⁹ Hammitt and Graham (1999) review over 25 contingent valuation studies where the size of the risk reduction was exogenously varied to the respondents. They find that the scope effect is more often satisfied when the size of the risk reduction is varied within a respondent (an internal test) than when it is varied across respondents (an external test), and that there are several studies where WTP does not vary systematically with the size of the risk reduction. They also report that in internal tests WTP is rarely found to be strictly proportional to the size of the risk reduction. In studies that conducted external scope tests WTP generally exhibits little responsiveness to the size of the risk reduction, and generally fails to be strictly proportional to the risk reduction.

expressed as a percentage of the baseline risk, this percentage being selected at random among a predetermined set of values (10%, 30% or 50%).

Respondents were next asked to assess their own risk of dying in a road-traffic accident (DEGRISK), and to value a risk reduction of 10%, 30%, 50% or 99% (RISKMD). Both risk reductions are private goods and are valid for one year only. In this paper, attention is restricted to the risks of dying in road accidents.

B. Hypotheses

We assume that:

$$(15) \quad WTP = \exp(\mathbf{x}_i \mathbf{b}_1) \cdot ABSRISK_i^{b_2} \cdot \exp(\mathbf{e}_i)$$

where \mathbf{x} is a $1 \times k$ vector of individual characteristics thought to influence risks, $ABSRISK$ is the absolute risk change ($ABSRISK = DEGRISK \times RISKMD$), and \mathbf{e} is an error term. On taking logs,

$$(16) \quad \log WTP = \mathbf{x}_i \mathbf{b}_1 + \mathbf{b}_2 \log ABSRISK_i + \mathbf{e}_i,$$

which can be re-written as:

$$(17) \quad \log WTP = \mathbf{x}_i \mathbf{b}_1 + \mathbf{b}_2 \log DEGRISK_i + \mathbf{b}_3 \log RISKMD_i + \mathbf{e}_i,$$

where $RISKMD$ is the percentage risk reduction, which is randomly assigned to the respondent in the survey, and is therefore considered an exogenous variable. If $\log DEGRISK$ and $\log WTP$ share common, unobservable respondent-specific factors, they are econometrically endogenous, which makes the OLS estimate of \mathbf{b}_2 biased and inconsistent. This problem can be addressed using instrumental-variable estimation techniques, such as two-stages least squares (2SLS).

In equation (17), the coefficients of $\log \text{DEGRISK}$ and $\log \text{RISKMD}$ are allowed to be potentially different from each other in order to test hypotheses about the determinants of WTP and about scope effects. Specifically, if $\mathbf{b}_2 = \mathbf{b}_3$, then respondents correctly valued the absolute risk reduction, which allows us to estimate the VSL. If $\mathbf{b}_2 = \mathbf{b}_3 = 1$, then WTP is strictly proportional to the size of the risk reduction, as economic theory suggests should be the case with small risks (Hammitt and Graham, 1999).

Should we find that $\mathbf{b}_2 \neq \mathbf{b}_3$, and that both \mathbf{b}_2 and \mathbf{b}_3 are different from zero, we would conclude that at least some weight has been given to the baseline risk (Baron, 1997; McDaniel, 1992). By contrast, if $\mathbf{b}_2 \neq \mathbf{b}_3$, $\mathbf{b}_2 = 0$ and $\mathbf{b}_3 \neq 0$, then WTP depends exclusively on the proportion, but not on the absolute risk reduction, and it is not possible to compute a meaningful VSL. One would expect the outcome of these tests of hypotheses to depend on whether baseline risks, DEGRISK , are treated as endogenous with WTP.

C. Endogeneity of Risks and WTP

If $\log \text{DEGRISK}$ is endogenous with $\log \text{WTP}$, we need an additional equation explaining $\log \text{DEGRISK}$:

$$(18) \quad \log \text{DEGRISK} = \mathbf{z}_i \mathbf{g}_1 + \mathbf{w}_i \mathbf{g}_2 + \mathbf{h}_i$$

where \mathbf{z} is a vector of instruments that overlap with some of the regressors \mathbf{x} in equation (15), \mathbf{w} is a vector of instruments excluded from the right-hand side of the WTP equation

to ensure identification, \mathbf{g}_1 and \mathbf{g}_2 are vectors of coefficients, and \mathbf{h}_i is an error term. ε and η are potentially correlated within a respondent.

We estimate the system of equations (17)-(18) by 2SLS using the broader of the two samples used by Persson et al. (2001).²⁰ The results of OLS estimation of (18) are reported in table 7 because they are of independent interest and because they are the first stage of the 2SLS procedure. Table 7 shows that subjective risks are related to age and age squared, and to miles driven per year. Other individual characteristics, such as gender, do not appreciably influence subjective risks.

As shown in table 7, we included in this regression various dummies, such as whether the respondent travels by bicycle, moped or motorcycle, and whether he wears a helmet when doing so, for two reasons. First, these variables are thought to influence the subjective chance of dying in a road traffic accident. Second, because they are excluded from the WTP equation, these variables provide identification restrictions for the 2SLS. However, table 7 shows that they are not significant predictors of subjective risks. We reach the same conclusion for the two education dummies (high school diploma and college degree), which were included in this equation for the same reasons.

Table 8 reports the results of OLS and 2SLS regressions of WTP on various regressors, including subjective risks. In addition to the risk variables, the regressors include the logarithmic transformations of income and age, the square of log age,²¹ log miles driven, a dummy indicating whether the respondent has ever been injured in an

²⁰ Specifically, following Persson et al., we form the sample by taking all observations with positive WTP, and replacing zero WTP values with a small positive number (i.e., WTP=2 SEK). Observations with missing baseline risk and/or missing WTP, baseline risk smaller than 1 in 100,000, WTP less than 1 and or WTP greater than 5% of annual income are excluded.

²¹ Including the logarithmic transformation of age and its square, rather than age and age squared, which were included in the first-stage regression, should help in identifying the coefficients of the WTP equation.

accident, and education and family composition dummies. As shown in table 8, regardless of the specification, the estimation technique used, and the restrictions placed on the parameters, WTP is well predicted by income, miles traveled, and the risk reduction. The effect of age is weak at best, and education and family status are not important.

In columns (A) and (B) of table 8, the restriction that $b_2 = b_3$ is imposed, which means that WTP is regressed on the size of the absolute risk reduction. Column (A) reports the results of OLS estimation that treats subjective risks, and hence the risk reduction, as exogenous, whereas column (B) reports the results of 2SLS estimation that treats subjective risk as endogenous with WTP *and* imposes the abovementioned restriction.

Column (A) shows that WTP does increase systematically with the size of the risk reduction, but in a less-than-proportional fashion: \hat{b}_2 is less than 1. The responsiveness of WTP to the size of the risk reduction is weak: doubling the size of the risk reduction would increase WTP by only 18%. With 2SLS (column (B)), one concludes that WTP is more responsive to the size of the risk reduction, since the 2SLS \hat{b}_2 is more than twice as large as its counterpart in column (A), implying that doubling the risk reduction increases WTP by 41%. The null hypothesis that absolute risk is exogenous is rejected soundly.

Columns (C) and (D) use OLS and 2SLS, respectively, but relax the restriction that $b_2 = b_3$. F tests of the null that these coefficients are equal reject the null at the conventional levels with OLS estimation, but fail to reject it with 2SLS. Inspection of the coefficient estimates and their t statistics, however, suggest that in both cases people were responding to the *proportions*, rather than the absolute risk reductions. This suggests that

with CV surveys about mortality risk reductions that present the risk reduction as a specified proportion of the baseline risk one should check whether subjects were truly responding to the absolute risk reduction, or simply to the percentage risk reduction, without considering the baseline. In testing these results, however, it is important to pay attention to the possible endogeneity of baseline risk with WTP, as conclusions may be sensitive to whether baseline risk is treated as endogenous with WTP.

Table 7. First-stage regression – Person et al. data.
Dependent variable: log DEGRISK. N=518.

	Coefficient	T statistic
Intercept	1.1095	1.85
Age	-0.0647**	-3.44
Age squared	0.00068**	3.16
Male	-0.0613	-0.69
Log km traveled in a car	0.2816**	3.83
Travels by moped or motorcycle (dummy)	0.0077	0.06
Travels by bicycle (dummy)	0.0735	0.42
Wears helmet when bicycling (dummy)	0.1875	1.36
Uses seatbelt when in back seat of car (dummy)	-0.1374	-1.28
High school diploma (dummy)	0.0489	0.39
College degree (dummy)	-0.0578	-0.46

** = significant at the 1% level.

Table 8. Second-stage results – Persson et al. data.
Dependent variable: log WTP. T stats in parentheses.

	A OLS n=676	B 2SLS N=579	C OLS n=676	D 2SLS N=579
Intercept	-1.0221 (-0.45)	-2.4185 (-0.97)	0.0810 (0.04)	-0.5672 (-0.18)
Log income per household member	0.4213** (2.47)	0.4777** (2.50)	0.3747* (2.19)	0.4373* (2.19)
Log km traveled in car	0.4949** (3.19)	0.4368** (2.60)	0.5502** (3.52)	0.6573** (2.47)
Log DEGRISK	0.1850** (2.32)	0.4092** (2.38)	0.0671 (0.73)	-0.6332 (-0.65)
Log RISKMD	0.1850** (2.32)	0.4092** (2.38)	0.5467* (3.35)	0.4420** (2.47)
Injured in accident (Dummy)	0.3779 (1.58)	0.4372^ (1.71)	0.3659 (1.53)	0.4262 (1.63)
Log age	-0.4943^ (-1.68)	-0.3308 (-1.03)	-0.5770* (-1.96)	-0.6910 (-1.48)
Log age squared	-0.0085 (-0.17)	-0.0105 (-0.18)	-0.0211 (-0.42)	0.0350 (0.45)
High school diploma	-0.1275 (-0.47)	0.0829 (0.28)	-0.0635 (-0.23)	0.1567 (0.50)
College degree	0.1298 (0.47)	0.3539 (1.16)	0.1627 (0.59)	0.3249 (1.03)
Household members ages 0-3	0.1754 (0.75)	0.1008 (0.42)	0.1441 (0.62)	0.0552 (0.22)
Household members ages 4-10	0.2401 (1.59)	0.3410* (2.14)	0.2396 (1.60)	0.3088^ (1.87)
Household members ages 11-17	-0.0445 (-0.26)	0.0366 (0.20)	-0.1594 (-0.09)	0.0269 (0.14)
Household members ages 18+	0.0543 (0.43)	0.1545 (1.14)	0.0372 (0.30)	0.1494 (1.07)
Test: $\beta_2=\beta_3$			F=6.44 Pval=0.011 4	F=1.18 Pval=0.277 6

Observations with missing baseline risk and missing WTP, observations with baseline risk smaller than 1 in 100,000, observations with WTP less than 1 and with WTP greater than 5% of annual income are excluded. Observations with WTP equal to zero are replaced by WTP=2.

VI. Internal Validity: The Relationship Between WTP and Income

Income is an important independent variable in regressions relating WTP to individual characteristics of the respondent. There are at least two reasons why researchers regress WTP on household (or personal) income. First, this is a common practice for testing the internal validity of the WTP responses, as theory suggests that WTP for mortality risk reductions should be positively associated with income. Second, there is much interest in the income elasticity of WTP for the purpose of predicting WTP at specified levels of income within the sample, or for benefit transfer purposes.^{22 23}

In many contingent valuation surveys about environmental quality or other public goods, researchers expect WTP to be a small fraction of the respondent's income. This expectation has led them, in some cases, to exclude from the sample respondents whose implied WTP is greater than, say, 5% of the respondent's income.

With reductions in one's own risk of dying, there is no particular reason to believe that WTP *should* be a small proportion of income. However, Persson et al. do omit from the usable sample respondents whose WTP exceeds 5% of household income, and Lanoie et al. (1995) find that their estimate of VSL for workers in the Montreal area, which is Can \$22-27 million, drops to Can \$15 million after excluding from the sample three

²² It is recognized, however, that knowing the income elasticity of WTP in a cross-sectional sample sense does not answer the important policy question of whether VSL should change over time, as income grows and the tradeoffs people are prepared to make between income and risk reductions change. Costa and Kahn (2002) circumvent this problem by estimating compensating wage studies for different years in the US. Using Census micro-data and fatality risk figures from the Bureau of Labor Statistics for 1940, 1950, 1960, and 1980, Costa and Kahn conclude that the quantity of safety has increased over time, and that the compensating differential has increased. The implied elasticity of VSL with respect to per capita GNP is 1.5 to 1.7. A meta-analysis of compensating wage studies by Viscusi and Aldy (2003) pegs the income elasticity of VSL to be 0.5-0.6, and certainly less than one. DeBlaeij et al. (2000) conduct a meta-analysis of the WTP to reduce transport risks, finding a considerable higher income elasticity of 1.33. Liu et al. (1997) compare estimates of the VSL from compensating wage studies in Taiwan based on 1982-1986 data with predictions based on VSL-income relationship from developing countries.

²³ Alberini (2004) discusses difficulties associated with measuring income in surveys.

workers whose WTP is one-third of their pre-tax income.²⁴ It seems, therefore, appropriate to check for respondents whose announced WTP is a relatively large proportion of income, and to examine how much the estimates of mean WTP and of the income elasticity of WTP change when these respondents are excluded from the sample. Respondents with very high announced WTP relative to income may, for example, have failed to give proper consideration to their budget constraint, have intentionally misrepresented their income, or have simply miscalculated it.

We use the data from two mortality risk surveys conducted in the US and Sweden, respectively—Alberini et al. (2004), and Persson et al. (2001)—to investigate these issues. We begin with the data from the Alberini et al. study. Results from estimating mean WTP after excluding respondents with WTP greater than a given percentage of income are shown in Table 9.²⁵ We vary this percentage from 25 (the least stringent criterion) to 2.5 (the most stringent criterion), showing how doing so excludes from 68 to 133 respondents (almost one-third of the sample). As shown in table 9, mean and median WTP do decline as we exclude more observations from the sample, but the change is within 10-12% of the original figures.

By contrast, what *does* change dramatically is the income elasticity of WTP, a key quantity when one wishes to (i) extrapolate study results to the general population, (ii) focus on the economically disadvantaged, and (iii) attempt benefit transfers to other countries or locales where income levels are different. As shown in table 9, the income

²⁴ Lanoie et al. (1995) value risk reductions in the car safety and workplace safety contexts. They survey employees of firms in the Montreal area, asking both willingness-to-pay and willingness-to-accept (WTA) questions. The Can \$22-27 million VSL figure refers to WTP for job safety (which is judged as more reliable than WTA).

²⁵ These estimates are based on a Weibull interval-data model that combines the responses to the initial WTP questions and to the follow-up questions in a double-bounded fashion.

elasticity of WTP is 0.16 when the full sample is used, 0.29 when persons whose implied WTP amount is greater than 10% of household income are excluded, 0.52 when we exclude persons whose WTP is greater than 5% of household income, and, finally, 0.92 when the most stringent criterion is used. Predictions for WTP as income changes would, therefore, vary dramatically, depending on which of these “cleaned” samples, and the corresponding income elasticity of WTP, one opts for.

Table 9. Outliers with respect to income.
Alberini et al. US Survey. WTP for 5 in 1000 risk reduction, wave 1, cleaned sample*

	Exclude if...	N	Mean WTP (\$)	Median WTP (\$)	Income elasticity of WTP
Least stringent	(all sample)	551	752.84 (88.37)	346.21 (28.45)	0.16
	WTP \geq 25% of household income	483	755.56 (90.84)	362.38 (31.97)	0.16
	WTP \geq 10% of household income	477	747.53 (90.02)	355.14 (29.24)	0.29
	WTP \geq 5% of household income	458	719.25 (89.21)	339.33 (30.02)	0.52
Most stringent	WTP \geq 2.5% of household income	418	678.39 (91.64)	302.26 (28.67)	0.92

* Excludes those respondents who failed the probability quiz and the probability choice.

Further investigation reveals that the 65 respondents who violated the most stringent exclusion criteria were slightly older than the remainder of the sample, but not significantly so (average ages were 57 and 54, t statistic of the null of no difference = 1.38), significantly less educated than the remainder of the sample (11.75 years of

schooling vs. 13.3, t statistic = -6.27), and reported much lower annual household income than the rest of the sample (sample averages: \$17,942 v. 56,151, t statistic -22.18).^{26 27} Moreover, they were twice as likely to indicate, in the debriefing section of the survey, that they had misunderstood the timing of the payment (27% of this group versus 13% of the remainder of the sample, t statistic = -4.66).

Results for the Persson et al. data are displayed in tables 10 and 11. One respondent in Persson et al.'s sample reports a WTP amount that is 83% of annual household income. Fortunately, the rest of the sample is more reasonable: Ninety-nine percent of the sample holds a WTP amount for reducing the risk of a fatal auto accident that is equal to or less than 12.5% of household income. In their analysis, Persson et al. discard from the usable sample observations such that WTP accounts for more than 5% of annual household income. This loses 29 observations.

Table 10 displays mean WTP for the full sample, and when persons with relatively high WTP/income ratios are excluded from the usable sample. This table shows that while median WTP remains the same for the various exclusion criteria, mean WTP jumps from 1875 to 2778 SEK when we reinstate into the sample those respondents whose WTP was more than 5% of household income. This is a 50% increase in WTP.

²⁶ The median annual household income is \$17,500 and \$55,000, respectively.

²⁷ It is possible that these respondents miscalculated or intentionally underreported their income. We regressed log income on age, age squared, education and the gender dummy for the full sample, and used the results of this regression to compute predicted income. For the 65 respondents with high WTP/income ratio, income predicted on the grounds of education, gender and age was always larger than reported income.

Table 10. Persson et al. study.
Effect of excluding observations with large annual WTP/household income ratios.
All values in 1998 SEK.

Exclusion criterion	Number of observations in the sample	Sample average WTP	Sample Median WTP
None	637	2778	1000
Respondents with zero income but positive WTP	637	2778	1000
WTP greater than 50% of household income (least stringent)	636	2635	1000
WTP greater than 25% of household income	633	2163	1000
WTP greater than 20% of household income	632	2151	1000
WTP greater than 12.5% of household income	631	2143	1000
WTP greater than 10% of household income	629	2134	1000
WTP greater than 5% of household income (most stringent)	618	1875	1000

Table 11 displays the income elasticity of WTP when observations where WTP accounts for a relatively large share of household income are omitted from the sample. As explained in section V, we estimate a system of simultaneous equations for log baseline risk and log WTP. The right-hand side of the WTP equation includes the logarithmic transformation of the absolute risk reduction, log miles traveled in a car in a year, a dummy accounting for previous injuries sustained in a car accident, log age, log age squared, two education dummies, and dummies for the size of the household in various age groups. Table 11 shows that income elasticity of WTP doubles when we move from the sample created with the least restrictive criterion to the most stringent criterion. It remains, however, relatively low (0.28).

We conclude that while there is no unambiguous criterion for considering one's WTP "large" relative to this person's income, researchers should experiment with checking how the estimates of WTP and other coefficients of interest are affected by including/excluding from the usable sample those respondents whose announced WTP is high relative to income. In the two examples presented in this section, doing so had a completely different impact on the estimates of mean WTP and on the income elasticity of WTP. These checks also pointed out that observations with a large WTP relative to income may be due to the respondent's failure to understand or retain details of the risk reduction scenario, as shown by the example based on the Alberini et al. data. They may also result from an inaccurate calculation or a deliberate misreporting of income on the part of the respondent.

Table 11. Persson et al. study.
Effect of excluding observations with large WTP/household income ratio on the income elasticity of WTP. 2SLS estimation, dependent variable: log WTP for risk of dying in a road traffic accident.

Exclusion criterion with respect to income	Number of observations	Income elasticity of WTP	Standard error
None	514	0.1475	0.1136
Respondents with zero personal income but positive WTP	514	0.1475	0.1136
WTP greater than 5% of household income	501	0.2850	0.1109
WTP greater than 10% of household income	509	0.2264	0.1139
WTP greater than 12.5% of household income	510	0.1937	0.1129
WTP greater than 20% of household income	511	0.1668	0.1126
WTP greater than 25% of household income	512	0.1418	0.1119
WTP greater than 50% of household income	514	0.1475	0.1136

Observations with missing baseline risk and missing WTP, observations with baseline risk smaller than 1 in 100,000, observations with WTP less than 1. Other regressors in the WTP equation: log degrid, log riskmd, log miles traveled in a car, previously injured in a traffic accident (dummy), log age, log age squared, two education dummies, dummies for household members. Coefficients of log degrid and log riskmd are restricted to be equal.

VII. Discussion and Conclusions

This paper examines the issue of the robustness of the estimates of the VSL from CV studies that elicit WTP for a reduction in the risk of dying. We illustrate the effects of maintained assumptions and techniques using the data from four recent CV surveys.

When dichotomous choice questions are used, we emphasize the importance of spreading bids nicely over a broad portion of the range of WTP (Cooper, 1993). We also show that the estimate of mean WTP from dichotomous-choice CV data can be extremely sensitive to the distributional assumption made by the researcher about the latent WTP

variable, and to the procedure used for calculating the welfare statistics. Median WTP tends to be less affected by these factors. We also find that claims about the shape of the relationship between WTP and certain covariates of interest (e.g., age) may no longer be valid when alternative distributional assumptions are made, alternative welfare statistics are used, or alternative (e.g., semiparametric) procedures are employed.

Outliers can, in general, be defined as observations that are distant from the remainder of the sample. How exactly “distant” is defined is, of course, a matter of interpretation. In this paper, we examine outliers in dichotomous choice CV samples in the sense of Copas (1988).

We also attempt to estimate discrete mixtures to identify yea-sayers, nay-sayers, and persons who answer the payment questions in a completely random fashion—another possible cause for outliers—but obtain implausible results that we attribute to the fact that the distribution assumed for WTP and/or the mixture we choose to work with fits the data poorly.

Next, we examine the internal validity of the WTP responses. Specifically, we look at the scope effect and at the association between WTP and income. Carson (2000) emphasizes that a CV study about mortality risk reductions must satisfy the “scope” test for the quality of the survey and its data to be considered acceptable. Testing for scope requires that the risk reduction be varied to the respondents by virtue of the experimental design. WTP will pass a “weak” scope test if it increases systematically with the size of the risk reduction. It will pass a “strong” scope test if it is proportional to the size of the risk reduction.

Testing for scope is straightforward when people are asked to value objective risk reductions which are varied to them in the study.²⁸ Testing for scope can be more complex when people are subjectively assessing their own baseline risks and/or risk reductions. This is because both of these variables may be affected by unobserved individual factors, which results in their econometric endogeneity. The Persson et al. survey is an example of one such study.

Using the Persson et al. data, we find that the sensitivity of WTP to the size of the risk reduction is indeed more pronounced when risk reductions meant as a specified proportion of a subjectively assessed baseline are treated as econometrically endogenous with WTP. We also show that with CV surveys conceived in this fashion it is important to test for whether people were truly valuing absolute risk reductions, or the mere proportion, without applying the latter to their baseline risks. In the Persson et al. study, people indeed appear to be responding to the percentage risk reduction, but not to the baseline risk, which makes it problematic to compute the VSL.

Turning to the other internal validity issue, we tackled the question whether one should omit observations with large WTP relative to income. We illustrate our checks using two applications where the omissions of these variables from the sample has the opposite effect on mean WTP and on the income elasticity of WTP, a key quantity when one wishes to (i) extrapolate study results to the general population, (ii) focus on the economically disadvantaged, and (iii) attempt benefit transfers to other countries or locales where income levels are different.

²⁸ Ideally, the scope tests should be external, i.e., performed using independent samples where respondents are faced with risk reductions of different size.

For the Alberini et al. application, further checks suggest that individuals with high announced WTP relative to income may have misunderstood some key aspects of the survey's risk reduction provision. This and the earlier checks imply that researchers should explore and report variants of their models driven by different assumptions, and attempt to recognize outliers and respondents who may have misunderstood some aspects of the valuation scenario.

References

- Alberini, Anna (1995), "Optimal Designs for Discrete Choice Contingent Valuation Surveys: Single-bound, Double-bound and Bivariate Models," *Journal of Environmental Economics and Management*, 28, 187-306.
- Alberini, Anna, and Richard T. Carson (2001), "Yea-Sayers, Nay-sayers, Or Just Plain Confused? Mixtures of Populations in Contingent Valuation Survey Responses," presented at annual meeting of the European Association of Environmental and Resource Economists, Southampton, England, June 2001.
- Alberini, Anna, Maureen Cropper, Alan Krupnick, and Nathalie Simon (2004), "Does the Value of a Statistical Life Vary with Age and Health Status? Evidence from the United States and Canada," *Journal of Environmental Economics and Management*, 48(1), 769-792.
- Alberini, Anna (2004), "Willingness to Pay for Mortality Risk Reductions: A Re-examination of the Literature," Final report to the US Environmental Protection Agency under Cooperative Agreement 015-29528, College Park, July.
- Atkinson, S.E., and Halvorsen, Robert (1990), "The Valuation of Risks to Life: Evidence from the Market for Automobiles," *The Review of Economics and Statistics*, 72(1), 133-136.
- Baron, Jonathan (1997), "Confusion of Relative and Absolute Risk in Valuation," *Journal of Risk and Uncertainty*, 14, 301-309.
- Blomquist, Glenn (1979), "Value of Life Saving: Implications of Consumption Activity," *Journal of Political Economy*, 96(4), 675-700.
- Cameron, Trudy A., and Michelle D. James (1987), "Efficient Estimation Methods for 'Closed-Ended' Contingent Valuation Surveys," *Review of Economics and Statistics*, 69(2), 269-276.
- Carson, Richard T. (2000), "Contingent Valuation: A User's Guide," *Environmental Science Technology*, 34, 1413-1418.
- Chen, Heng Z. and Alan Randall (1998), "Semi-nonparametric Estimation of Binary Response Models with an Application to Natural Resource Valuation," *Journal of Econometrics*, 76, 323-340.
- Collett, D. (1991), *Modelling Binary Data*, London, UK: Chapman and Hall.

- Cooper, Joseph A. (2002), "Flexible Functional Form Estimation of Willingness to Pay Using Dichotomous Choice Data," *Journal of Environmental Economics and Management*, 43, 267-279.
- Cooper, J. (1993), "Optimal Bid Selection for Dichotomous Choice Contingent Valuation Surveys," *Journal of Environmental Economics and Management*, 24, 25-40.
- Copas, J. B. (1988), "Binary Regression Models for Contaminated Data," *Journal of the Royal Statistical Society, Series B*, 50(2), 225-265.
- Costa, Dora L. and Matthew E. Kahn (2002), "Changes in the Value of Life, 1940-1980," NBER Working Paper 9396, Cambridge, December.
- Creel, Michael and John B. Loomis (1997), "Semi-nonparametric Distribution-Free Dichotomous Choice Contingent Valuation," *Journal of Environmental Economics and Management*, 32, 341-358.
- Crooker, John R and Joseph A. Herriges (2004), "Parametric and Semi-nonparametric Estimation of Willingness-to-Pay in the Dichotomous Choice Contingent Valuation Framework," *Environmental and Resource Economics*, 27, 451-480.
- deBlaey, Arianne, R.J.G. Florax, and P. Rietveld (2000), "The Value of Statistical Life in Road Safety: A Meta-analysis," Working paper TI 2000-089/3, Tienbergen Institute, Amsterdam.
- Gerking, Shelby, Menno De Haan and William Schulze (1988), "The Marginal Value of Job Safety: A Contingent Valuation Study," *Journal of Risk and Uncertainty* 1(2): 185-200.
- Hammitt, James K. and John D. Graham (1999). "Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability?" *Journal of Risk and Uncertainty*, 8, 33-62.
- Hoehn, John P. and Alan Randall (1987), "A Satisfactory Benefit Cost Indicator from Contingent Valuation," *Journal of Environmental Economics and Management*, 14, 226-247.
- Johannesson, Magnus, Per-Olov Johansson, and Karl-Gustav Löfgren (1997), "On the Value of Changes in Life Expectancy: Blips Versus Parametric Changes," *Journal of Risk and Uncertainty*, 15, 221-239.
- Johannesson, Magnus, Bengt Jonsson, and Lars Borquist (1991), "Willingness to Pay for Antihypertensive Therapy—Results for a Swedish Pilot Study," *Journal of Health Economics*, 10, 461-474.

- Kanninen, Barbara J. (1993), "Optimal Experimental Design for Double-bounded Dichotomous Choice Contingent Valuation," *Land Economics*, 69(2), 138-146.
- Kanninen, Barbara J. (1995), "Bias in Discrete Response Contingent Valuation," *Journal of Environmental Economics and Management*, 28, 114-125.
- Krupnick, Alan, Anna Alberini, Maureen Cropper, Nathalie Simon, Bernie O'Brien, Ron Goeree, and Martin Heintzelman (2002), "Age, Health, and the Willingness to Pay for Mortality Risk Reductions: A Contingent Valuation Survey of Ontario Residents," *Journal of Risk and Uncertainty*, 24, 161-184.
- Lanoie, Paul, Carmen Pedro and Robert Latour (1995), "The Value of a Statistical Life: A Comparison of Two Approaches," *Journal of Risk and Uncertainty*, 10, 235-257.
- Liu, Jin-Tan, James K. Hammitt and Jon-Long Liu (1997), "Estimated Hedonic Wage Function and Value of Life in a Developing Country," *Economic Letters*, 57, 353-358.
- McDaniel, Timothy (1992), "Reference Points, Loss Aversion, and Contingent Values for Auto Safety," *Journal of Risk and Uncertainty*, 5, 187-200.
- Persson, Ulf, Anna Norinder, Krister Halte, Katarina Gralen (2001), "The Value of a Statistical Life in Transport: Findings from a New Contingent Valuation Study in Sweden," *Journal of Risk and Uncertainty*, 23(2), 121-134.
- Pope, C. Arden III, et al. (1995), "Particulate Air Pollution as a Predictor of Mortality in a Prospective Study of US Adults," *American Journal of Respiratory Critical Care Medicine*, 151, 669-674.
- Skrzycki, Cindy (2003), "Under Fire, the EPA Drops the 'Senior Death Discount'," *Washington Post*, May 13.
- Viscusi, W. Kip and Joseph E. Aldy (2003), "The Value of a Statistical Life: A Critical Review of Market Estimates throughout the World," *Journal of Risk and Uncertainty*, 27(1), 5-76.