Eliciting Subjective Beliefs About Mortality Risk Orderings

by

Glenn W. Harrison & Elisabet E. Rutström †

April 2005

ABSTRACT

We develop an experimental method to elicit subjective beliefs about the ordering of mortality risk over different causes of death. The experimental procedure emphasizes incentive-compatibility, so that the individual has a positive financial incentive to respond truthfully. We also consider the extent to which individuals have subjective beliefs for sub-segments of the population that are more accurate than their beliefs about the risks for the population as a whole. We propose several hypotheses concerning the degree of familiarity of the risks, and find that the evidence supports those hypotheses. The evidence also suggests that there is no discernible difference between beliefs elicited using hypothetical or real financial rewards in the elicitation format we use. Our findings restore some confidence in the ability to elicit beliefs about mortality risks, and therefore to get reliable estimates of the monetary value of a statistical life.

† Department of Economics, College of Business Administration, University of Central Florida (Gharrison@bus.ucf.edu and Erutstrom@bus.ucf.edu). Both authors are also affiliated with the Institute of Simulation & Training, University of Central Florida. Rutström thanks the U.S. National Science Foundation for research support under grants NSF/IIS 9817518, NSF/MRI 9871019 and NSF/POWRE 9973669. We are grateful to Maribeth Coller, Shelby Gerking, Melayne McInnes and two referees for comments. Supporting data, computer code and instructions are stored in project "Eliciting Mortality Risk Orderings" at the ExLab Digital Archive located at http://exlab.bus.ucf.edu. We dedicate this paper to the memory of J. Clay Lesley, who assisted us in the design and execution of the experiments and was a valued student, colleague and friend.

Reliable measures of subjectively perceived mortality risks are important to the accurate assessment of environmental costs and benefits. Nevertheless, as has been pointed out by Viscusi [1993], such measures have hereto forth been biased and unreliable. We show that certain types of subjectively perceived mortality risk measures are more reliable than others, restoring somewhat the confidence in our ability to perform accurate environmental cost-benefit analyses.

The monetary value of a statistical life (VSL) has been an important input into environmental cost-benefit analysis. Many environmental regulations entail some expectation of a reduction in premature mortality. This reduction in mortality constitutes a major benefit from environmental policies. One important example is the retrospective U.S. Environmental Protection Agency (USEPA) [1997] cost-benefit analysis of the provisions of the Clean Air Act, which concluded that mortality reduction benefits between 1970 and 1990 represented over $16 trillion in present value terms. Uncertainty over the VSL translates directly into uncertainty over the benefits of environmental policies, such as those embodied in the Clean Air Act. The USEPA [1997; Figure I-2, p. I-21]) calculates that benefits could be as low as $5.6 trillion or as high as $49.4 trillion, using the 5th and 95th percentiles, and that "most of the uncertainty in the total estimated benefit levels comes from uncertainty in the estimate of the economic valuation of mortality." In fact, the 5th and 95th percentiles of total benefits are $7 trillion and $45 trillion if one *only* allows the VSL estimate to exhibit uncertainty.

VSL estimates have typically been inferred from labor market choices that workers make over alternative risky jobs, from the prices consumers have paid for goods with negative environmental or health effects, or from direct surveys of willingness to pay to avoid the risk of death. In the first two cases the wages or prices are assumed to include a risk premium, reflecting the VSL. Estimates of these risk premia rely on the use of reliable measures for the *subjectively perceived* *risk* as explanatory variables. These calculations are conceptually straightforward. In practice, however, it has been hard to find ideal measures for the subjectively perceived risk; the key explanatory variable needed to make the simple logic operational.[1]

Lacking measures of the subjective risk beliefs, investigators have resorted to the use of objective measures as proxies for the subjective ones. Even these are problematic, often due to the lack of completeness and accuracy of the field data used. Inaccuracy is often caused by excessive aggregation of the classifications of risky activities. Data are incomplete because of a lack of measures of latent risk. All existing studies appropriately caution that the empirical application of one proxy or another proxy can often change the estimated value dramatically (e.g., Viscusi [1993]). Perhaps the most notable example of this is the use of one major U.S. governmental survey of risk in place of another survey by Moore and Viscusi [1988], leading to a statistically significant change in the VSL estimate from $2.5 million to $7.3 million in 1990 dollars. Indeed, the title of their study was: "Doubling the Estimated Value of Life: Results Using New Occupational Fatality Data."

The desire to avoid using proxies for the subjectively perceived risk has led to attempts to directly elicit risk perceptions in controlled laboratory experiments. The subjectively perceived risks elicited in laboratory settings have, however, been suspect. For example, since Lichtenstein et al. [1978], it has been generally assumed that people systematically overestimate the frequency of likely

[1] The basic idea underlying the VSL calculation is that the observed labor market choices of workers will reflect their willingness to accept risks, just as it reflects other aspects of the job. All other things being equal, a worker that is risk averse would require a higher wage to work in a job that is riskier than some alternative job. Thus one would expect that the different wages observed for different jobs should be positively correlated with the riskiness of the job. If so, one can back out a measure of the extra dollars per year a typical worker requires to accept a job that has an extra 1-in-10,000 chance of being associated with a fatal injury. If this wage premium, or risk premium, is $500, then we can multiply by 10,000 to ascertain the value that is *implicitly* placed on a *statistically* certain death. In this case the value would be $500 × 10,000 = $5 million, which happens to be a good approximation of the ball-park of many of the estimates (e.g., Viscusi [1993; Table 2]).

events and underestimate the frequency of unlikely events.[2]

We develop an instrument which can be used to elicit subjective *rankings* of mortality risk, as a first step in an effort to develop more reliable measures of risks as perceived by individuals. Since information about the level of the risk subsumes information about the ordering, but not vice versa, one expects the ordering task to be less costly in terms of its information requirements than the level determination task. We therefore maintain that subjects hold more accurate perceptions of the former than of the latter, at least for some risks.[3] We use this instrument to test several hypotheses regarding the ability of subjects to report risks accurately.

First, we expect that subjects might have relatively good estimates of the ranks of those mortality risks with which they are more familiar, in the sense of having observed others being exposed to them, than of those with which they are less familiar. Familiarity through such observations is more likely the higher the frequency of occurrence of the causes of deaths in the general population, since it makes them more visible, and thus more observable. We therefore propose the hypothesis that subjects are better able to rank those mortality risks which occur with a higher frequency in the general population than those which occur rarely.[4]

Alternatively, subjects may have more information and knowledge about the ranking of mortality risks for their own age group than for other age groups. Such own age group mortality

---

[2] Morgan and Henrion [1990; ch.6] review the early, extensive literature.

[3] One policy concern that motivates this hypothesis is the importance of beliefs about risks that will affect the individual well into the future. For example, most smokers begin smoking at a very young age, around 13 or 14 if surveys are to be believed, but the vast bulk of health effects of smoking do not occur until much later in life. If smoking is completely addictive, the only relevant risk perception is the one that beginning smokers hold, not the one that adult smokers hold. Of course, smoking is not completely addictive, but the same logic applies even when quitting is a difficult process that entails short-term costs. Evidence from Viscusi [1990] that mature smokers over-estimate the mortality risk of smoking says nothing about the risk perceptions relevant to the onset of smoking by much younger individuals.

[4] This hypothesis contrasts with the popular view since Lichtenstein et al. [1978], which claims that likely events are underestimated, and that they are estimated less accurately than somewhat less likely events.

risks are arguably more relevant to the subjects, and hence they may possess more information about them (Benjamin and Dougan [1997]). Extending this hypothesis, subjects may have more knowledge of mortality risks for age groups which they were once a member of.[5]

Finally, we seek to test the extent to which hypothetical surveys will elicit responses that differ from surveys in which subjects are financially rewarded for their accuracy. Many field surveys are non-salient, and it is important to identify if there are any systematic differences in elicited ranks when there is a positive incentive for accuracy. It is commonly observed that the lack of incentives for accurate reporting leads not only to higher variances but also to significant biases in responses. To take one example, consider the second set of experiments conducted by Neill et al. [1994] using open-ended elicitation for private goods with Vickrey auctions. They found mean valuation of $12 and $301 using real and hypothetical rewards, respectively, and standard deviations of $21 and $590.[6] One attractive feature of our instrument is that the financial incentive for accurate responses is transparent and strong, which is not always the case with proper scoring rules used to elicit probabilities of risk (e.g., von Winterfeldt and Edwards [1986; Chapter 11]).[7]

---

[5] This hypothesis could be extended to other demographic characteristics that are relevant to health conditions, as well. In this study we restrict our attention to age, since that is a variable for which health conditions and disease vary significantly. We concede that the age variation in our convenience sample of college students may not be as great as one would expect in the general population, but we do have reasonable variation in age: 31% of our sample was 19, 13% was 20, 13% was 21, 12% was 22, 5% was 23, 5% was 24, each of the ages 25 to 31 had at least 2.2% of the sample, and we had 5 subjects spread between ages 32 and 49.

[6] These differences are reduced when one trims larger hypothetical values, but a significant difference persists. Moreover, trimming away larger hypothetical values is simply assuming the bias away. Similar examples can be readily found using other elicitation formats. In the context of dichotomous-choice elicitation methods, see Cummings, Harrison and Rutström [1995]; in the context of public referendum elicitation methods, see Cummings et al. [1997]; and in the context of posted-price elicitation methods, see Coller and Williams [1999]. Harrison and Rutström [2005] review the general experimental literature on hypothetical bias.

[7] It has been known for some time that subjective probabilities can be elicited in a manner that offers incentives for truthful revelation (e.g., Savage [1971]). Thus it is possible to turn an unobservable determinant of behavior into an observable. Unfortunately, the standard implementation of these methods for eliciting beliefs suffer from some serious design defects. The most important defect is that rewards for truthful

In section 1 we introduce our instrument for eliciting the ranks of mortality risks, and state the research hypotheses we propose to test. In section 2 we present the results from experiments in which subjects were financially rewarded for accuracy. In section 3 we modify the instrument to examine non-salient, hypothetical responses. In section 4 we draw some general conclusions.

Our results are encouraging, in the sense that we find that subjects do seem to have more accurate orderings over mortality risks when they occur more frequently in the general population than not, and also over those for their own age group and for the age group immediately preceding their own rather than older age groups. We also find, quite surprisingly, that hypothetical responses in this format are not significantly different from those of financially motivated subjects.

1. Experimental Design

The first general feature of the new survey instrument is a recognition of the perceptual difficulty of certain tasks, such as the reporting of open-ended valuation and frequency questions. The responses here are elicited in a manner which restricts the numerical range in which subjects have to provide estimates of the variables in question. In particular, we elicit reports about the ordering rather than the level of risks. Such restrictions are assumed to simplify the problem to the subjects, compared to more open-ended frequency or valuation questions, since information about the frequency of events subsumes information about the ordering, but not vice versa.

The second feature of the new instrument is that risks are separately elicited for age groups that match those of the individual respondents and for other age groups. Thus we should be able to

---

reporting are dismal with the usual (quadratic) scoring rules. This does not mean that the reports themselves are poor, but it does mean that it is hard to maintain the illusion that they have been elicited in a controlled manner. Good examples of the application of these belief elicitation approaches in experimental economics include McKelvey and Page [1990], Croson [2000], Dufwenberg and Gneezy [2000], McDaniel and Rutström [2001] and Nyarko and Schotter [2002].

determine if responses are more accurate for the subject's own age group, than for the population as a whole, as hypothesized in Benjamin and Dougan [1997].

The third feature is the use of simple financial incentives to elicit true responses. This feature is explained below, and includes an evaluation of the presence of hypothetical bias when an alternative version is presented to subjects in a non-salient manner.

The survey asked subjects to say what they believed to be the *rank* of each of 12 causes of death in the United States in 1995. These are the 12 major causes of death for the adult population in that year, as defined in official government statistics (discussed below). The subject was asked to rank these causes of death for each of four age groups, so the subject was asked to provide 48 rankings in all.

The factual data were obtained from the U.S. Department of Health publication *Health United States 1986-96*, Table 34, for the 10 main mortality causes in each age group. Details for the remaining two causes in each age group were obtained from the CDC[8] in 5-year intervals and aggregated to the age intervals used here.

The motivation for this design is a concern that the well-known elicitation biases for mortality risk reported by Lichtenstein et al. [1978] may be due to the task being difficult for the subjects. Each subject in their survey was told that roughly 50,000 people died in 1978 from traffic accidents, and asked to state their belief about how many died from the other 40 conditions listed. Apart from the difficulty of the subjects having to report beliefs over the number of deaths, they were asked to report beliefs for the general population. As noted by Benjamin and Dougan [1997], this may be a difficult task for subjects who have relatively little interest in knowing the risks facing the "rest of the population." Benjamin, Dougan and Buschena [2001] replicate the original

---

[8] From http://www.cdc.gov/nchs/datawh/statab/unpubd/mortabs/gmwki.htm.

Lichtenstein et al. [1978] survey, in addition to testing for differences in responses based on own vs. other age groups. Neither of these studies employed financial incentives.

B. Survey Instrument

The survey instrument was administered in writing. The text is brief, and the version with salient financial rewards is reproduced below in full. We add markings to indicate which text was used in the salient version with real financial rewards for accuracy, and which text was used in the non-salient version :

RANKING THE CAUSES OF DEATH

In 1995 there were 2,312,132 recorded deaths in the United States. Of these, 34,244 or 1.5% were in the 15-24 age group; 160,015 or 6.9% were in the 25-44 age group; 378,512 or 16.4% were in the 45-64 age group; and 1,694,326 or 73.3% were 65 or over. A number of causes of death are listed below. These are the classifications used by the United States Department of Health. The examples are provided by us to help you understand what type of disease or condition is referred to.

Please tell us what rank you think these causes of death have for each age group, where the ranking refers to the frequency of each cause of death in 1995. You should write the number 1 for the cause of death that you believe was most common in 1995 for that age group. Then write the number 2 for the cause of death you believe to be the next most common for that age group, and so on. Write the number 12 for the cause of death you believe to the least common for that age group.

[SALIENT VERSION ONLY: You will be paid according to how well you rank these, compared to the official tabulations put out by the U.S. Department of Health. For every rank that you get correct you will be paid $1. For every rank that you are incorrect, your payoff will decrease by 10 cents for each rank that you were off. Your personal payoff does not depend on what other people in the room write as their rankings.

Consider some examples. If you rank some cause of death as #2 and it is actually #5, your payoff for that ranking will be $1.00 minus $0.30, where the 30 cent deduction is 10 cents times the 3 ranks you were off (since 3 = 5 - 2). Thus you would be paid 70 cents for this ranking. If you rank some cause of death as #10 and it is #11 you would receive $1.00 minus $0.10, or 90 cents, for that ranking. If you rank some cause of death as #7 and it is #7 you would receive $1.00 for that ranking. Finally, if you rank some cause of death as #12 and it is #1, you would lose 10 cents on that ranking.

Your payoff is based on your responses for <u>all</u> causes of death and <u>all</u> age groups, so you will maximize the amount of money you earn by getting every rank as close to the true rankings as you can. In the above example, the person would have earned $2.50 (= 70 cents + 90 cents + $1.00 - 10 cents) for the four rankings we went through. Each person has 48 rankings to fill in below and 48 chances to earn money. If you get each and every ranking correct you would therefore earn $48.

If you leave one ranking missing you will not be paid, so check when you are finished that all cells in the table below are filled in.]

[NON-SALIENT VERSION ONLY: You will be paid $10 for your time. We would like you to try to rank these as accurately as you can, compared to the official tabulations put out by the U.S. Department of Health.

When you have finished please check that all cells in the table below are filled in.]

If you do not know what conditions or diseases the official cause of death refers to you should make your best estimate.

PLEASE STATE YOUR RANKINGS FOR EACH AGE GROUP

Rank the most common cause of death as 1, the next most common as 2, and so on.

There are 12 causes of death, so be sure to fill in 12 ranks for each age group.

Cause of Death, in alphabetical order	Age 15-24	Age 25-44	Age 45-64	Age 65 and over
Alzheimer's disease				
Chronic liver disease and cirrhosis				
Diabetes mellitus				
Diseases of the heart (e.g., heart attack)				
Cerebrovascular diseases (e.g., stroke)				
Chronic obstructive pulmonary diseases (e.g., bronchitis, emphysema, asthma)				
HIV infection				
Homicide and legal intervention				
Malignant neoplasms (cancers)				
Pneumonia and influenza				
Suicide				
Unintentional injuries (e.g., car accidents)				

C. Research Hypotheses

All our hypotheses are formulated in terms of ranking errors. These ranking errors are calculated as the absolute value of the difference between the stated rank and the true rank. Thus, if a subject states that some cause of death is ranked 8th and it is actually ranked 6th or 10th, the error would be 2. These errors therefore take on integer values for each subject, between 0 and 12 for each of the 12 responses in each age bracket. Since each subject is asked to state rankings for four age brackets, there are 48 responses from each subject and 48 "error responses" from each subject.

The first hypothesis to be tested is that the subject will have more accurate beliefs over risks pertaining to their own rather than others' age group , all other things being equal. This means that the subject is predicted to make fewer errors for the causes of death in his own age group. Consider a dummy variable that takes on the value 1 for the twelve responses that each subject gives (out of 48 in all) that refer to his own age group, and takes on the value 0 for the thirty-six responses that refer to other age groups. We therefore predict that the coefficient on this dummy variable will be negative and statistically significant, implying that subjects make *fewer errors* when responding about causes of death in their own age group.

The second hypothesis, which is an extension of the first hypothesis in some respects, is that subjects will have better beliefs of risks pertaining to people in age groups they have lived through or are living through now. In other words, even if the subjects are in the 25-44 age group, they are hypothesized to have better information about the 15-24 age group since they have experienced it. Thus we define a dummy variable to be equal to 1 for all responses that a given subject provides that pertain to causes of death in age groups *strictly prior to* the age group they fall within. For subjects in the youngest age group this dummy variable is equal to 0 for all responses, but for all other (older) subjects it will detect this "cumulative" age effect. Again, the coefficient is predicted to be negative and statistically significant.

The third hypothesis is that subjects will have better estimates of those risks which they observe more often because they occur more frequently in the general population. Thus it is more likely that an individual would have relatively poor knowledge of causes of death that only kill a handful of people each year, unless they are "exotic" or politically noteworthy for other reasons (e.g., the anthrax deaths of 2001 in the United States). Since individuals are simply exposed to more deaths that occur more often in the population, one would expect the subject to have better knowledge of them. This hypothesis can be captured by introducing a variable equal to the true population incidence of the cause of death. Specifically, the variable is set equal to the proportion of the 48 deaths that was observed in 1995. The hypothesis predicts that the coefficient on this variable will be negative and statistically significant.

The fourth hypothesis is that financial incentives for accurate reporting will have the effect of significantly reducing the bias *and variance* of responses. A simple dummy variable captures this treatment condition. A large literature, particularly in the elicitation of values for environmental goods, shows significant biases and variances in hypothetical responses (see Harrison [2004] for a review).

It is likely that the ranking task is more difficult when the absolute number of deaths due to some cause is very close to the absolute number of deaths of some other cause of death in the same age group. That is, the difference in the number of deaths between two cells corresponds to a perceptually difficult task when the subject is asked to differentiate them by assigning distinct ranks. For example, even if the subject correctly knows that two causes of death are ranked #7 and #8 in relation to the other 10 causes of death, she may not be able to easily assign one to the 7^{th} rank and the other to the 8^{th} rank. In the extreme case, two causes of death might differ in only one death (as a factual matter, no two causes of death had the same number of deaths in any given age bracket). We test our hypotheses on samples of responses that include such perceptually difficult tasks and on samples that do not.

To make this sampling method operational a threshold of perceptive difficulty must be established. This turns out to be relatively straightforward, although there are different ways to operationalize it. Consider the data in Table 1. The top panel lists the 12 causes of deaths for the youngest age group, and the next three panels do likewise for the other age groups. Focus on the top panel to see the general logic. The column marked "Actual number of deaths" lists the actual number of deaths, by which the rows in the panel are sorted. The percentage of deaths is recorded in the column marked "Percent of deaths," where this refers to the percentage of deaths from the 12 causes of death listed here (there are other causes of death apart from these 12, of course). Finally, the column marked "Difference in sorted % of deaths compared to the prior (sorted) cause" shows the simple difference between the percentage of that (sorted) cause of death and the percentage of the next less likely cause of death. Thus the first sorted cause of death, Alzheimer's disease, must have a 0% here since there is no less likely cause of death in this age group. The second sorted cause of death is Chronic liver disease and cirrhosis, and it has a percentage of 0.11% compared to the Alzheimer's percentage of 0%; thus the difference between the two is 0.11% = 0.11% - 0%.

Using these results we can quickly identify cells that are perceptually difficult to differentiate from other cells. These are marked in the far right column. One asterisk indicates those cells with a 1% or less difference between the percentage of the closest (lower) cell.[9] Two asterisks indicate those cells with a 2% or less difference. We use the 2% indicators to generate dummy variables to

[9] One weakness of this metric of perceptual difficulty is that it only compares the fraction of deaths to those in the "prior" cell, where priority is defined in terms of size. In this sense it is "backward looking," and ignores perceptual difficulty for a given cell in relation to those after it. Nonetheless, there is something naturally attractive about using a measure of perceptual difficulty that is computationally simple.

Table 1: Identifying Perceptually Difficult Cells in the Overall Elicitation Task

True number of deaths (sorted)	Percent of deaths (sorted)	Name of cause of death (sorted)	Difference in sorted % of deaths compared to the prior (sorted) cause	Tough Tasks
		A. AGES 15-24		
0	0.00%	Alzheimer's disease	0.0%	*
33	0.11%	Chronic liver disease & cirrhosis	0.1%	*
136	0.45%	Diabetes mellitus	0.3%	*
172	0.57%	Cerebrovascular diseases	0.1%	*
207	0.69%	Pneumonia and influenza	0.1%	*
246	0.82%	COPD	0.1%	*
629	2.10%	HIV infection	1.3%	**
1039	3.46%	Diseases of the heart	1.4%	**
1624	5.47%	Malignant neoplasms	2.0%	
4784	15.94%	Suicide	10.5%	
7284	24.27%	Homicide and legal intervention	8.3%	
13842	46.12%	Unintentional injuries	21.8%	
30014				
		B. AGES 25-44		
7	0.00%	Alzheimer's disease	0.0%	*
1205	0.84%	COPD	0.8%	*
2102	1.46%	Pneumonia and influenza	0.6%	*
2458	1.71%	Diabetes mellitus	0.2%	*
3492	2.42%	Cerebrovascular diseases	0.7%	*
4309	2.99%	Chronic liver disease & cirrhosis	0.6%	*
10280	7.14%	Homicide and legal intervention	4.1%	
12759	8.86%	Suicide	1.7%	**
17064	11.84%	Diseases of the heart	3.0%	
21985	15.26%	Malignant neoplasms	3.4%	
30754	21.35%	HIV infection	6.1%	
37660	26.14%	Unintentional injuries	4.8%	
144075				

C. AGES 45-64

367	0.11%	Alzheimer's disease	0.0%	*
2879	0.88%	Homicide and legal intervention	0.8%	*
5537	1.69%	Pneumonia and influenza	0.8%	*
7336	2.24%	Suicide	0.5%	*
10499	3.20%	HIV infection	1.0%	*
10603	3.23%	Chronic liver disease & cirrhosis	0.0%	*
12184	3.71%	Diabetes mellitus	0.5%	*
12744	3.88%	COPD	0.2%	*
15208	4.63%	Cerebrovascular diseases	0.8%	*
16004	4.88%	Unintentional injuries	0.2%	*
102738	31.31%	Diseases of the heart	26.4%	
132084	40.25%	Malignant neoplasms	8.9%	
328183				

D. AGES 65 AND OVER

769	0.05%	HIV infection	0.0%	*
1084	0.08%	Homicide and legal intervention	0.0%	*
5849	0.41%	Suicide	0.3%	*
10232	0.73%	Chronic liver disease & cirrhosis	0.3%	*
20230	1.43%	Alzheimer's disease	0.7%	*
29099	2.06%	Unintentional injuries	0.6%	*
44452	3.15%	Diabetes mellitus	1.1%	**
74397	5.28%	Pneumonia and influenza	2.1%	
88478	6.28%	COPD	1.0%	**
138762	9.84%	Cerebrovascular diseases	3.6%	
381142	27.03%	Malignant neoplasms	17.2%	
615426	43.65%	Diseases of the heart	16.6%	
1409920				

identify the tough cells, therefore including all the cells with either one or two asterisks. We expect the main hypothesis tests to have higher power when applied to the sub-sample that does not include the perceptually difficult tasks.

2. Experimental Results with Monetary Incentives

Subjects for the treatment with monetary incentives were recruited from an advanced undergraduate class in Environmental Economics at the Moore School of Business in late September 2001. Forty-five subjects participated, and the entire experiment lasted roughly 20 minutes.

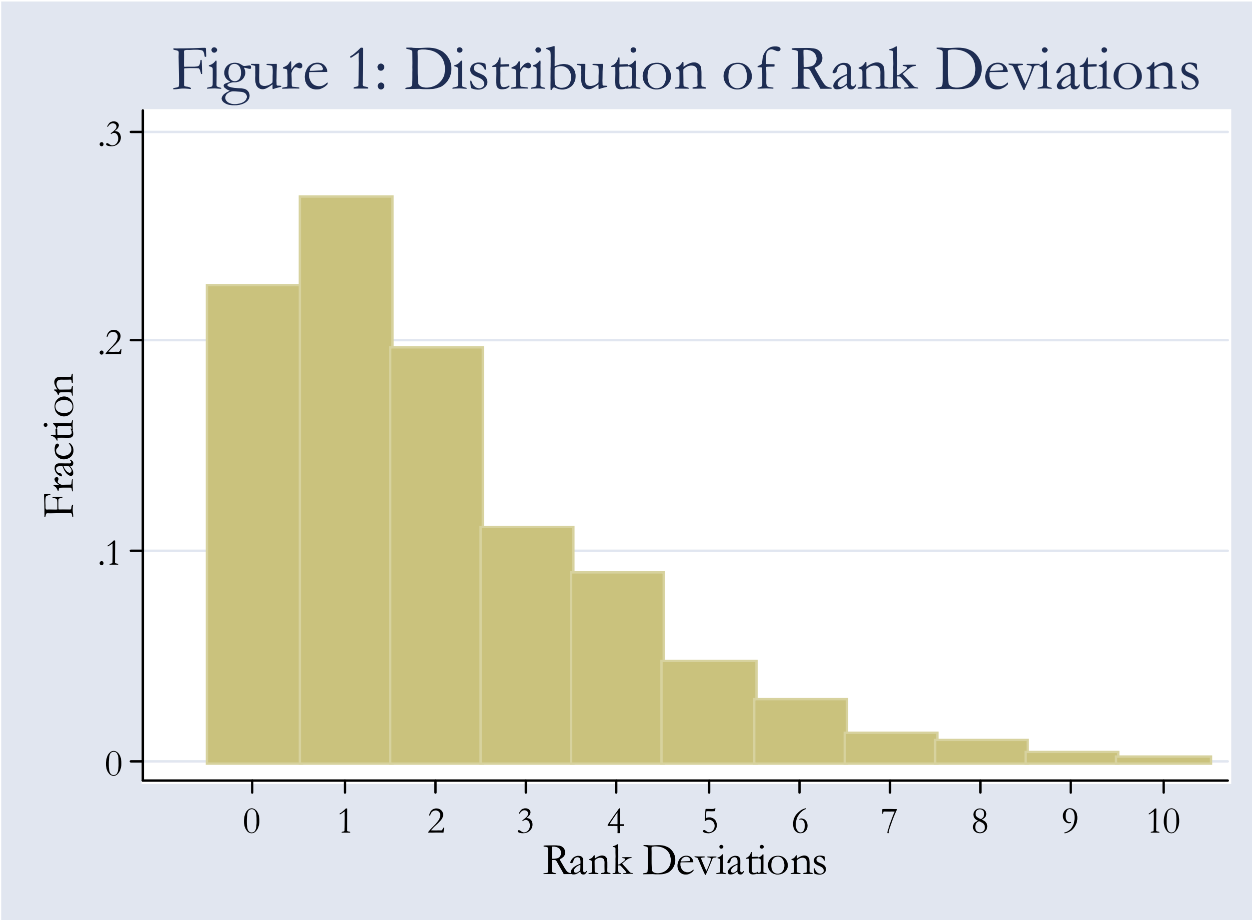
Table 2 displays the mean and standard deviation of the main variable of interest here, the absolute value of deviations in the actual from reported ranks, for the treatment with monetary incentives.[10] These data are a panel, since there are 48 responses from each individual. Hence it is appropriate to report the standard deviation "within individuals" as well as "between individuals," along with the overall standard deviation.

Table 2: Descriptive Statistics for Rank Deviations

Sample size = 2160, consisting of 48 decisions by 45 subjects

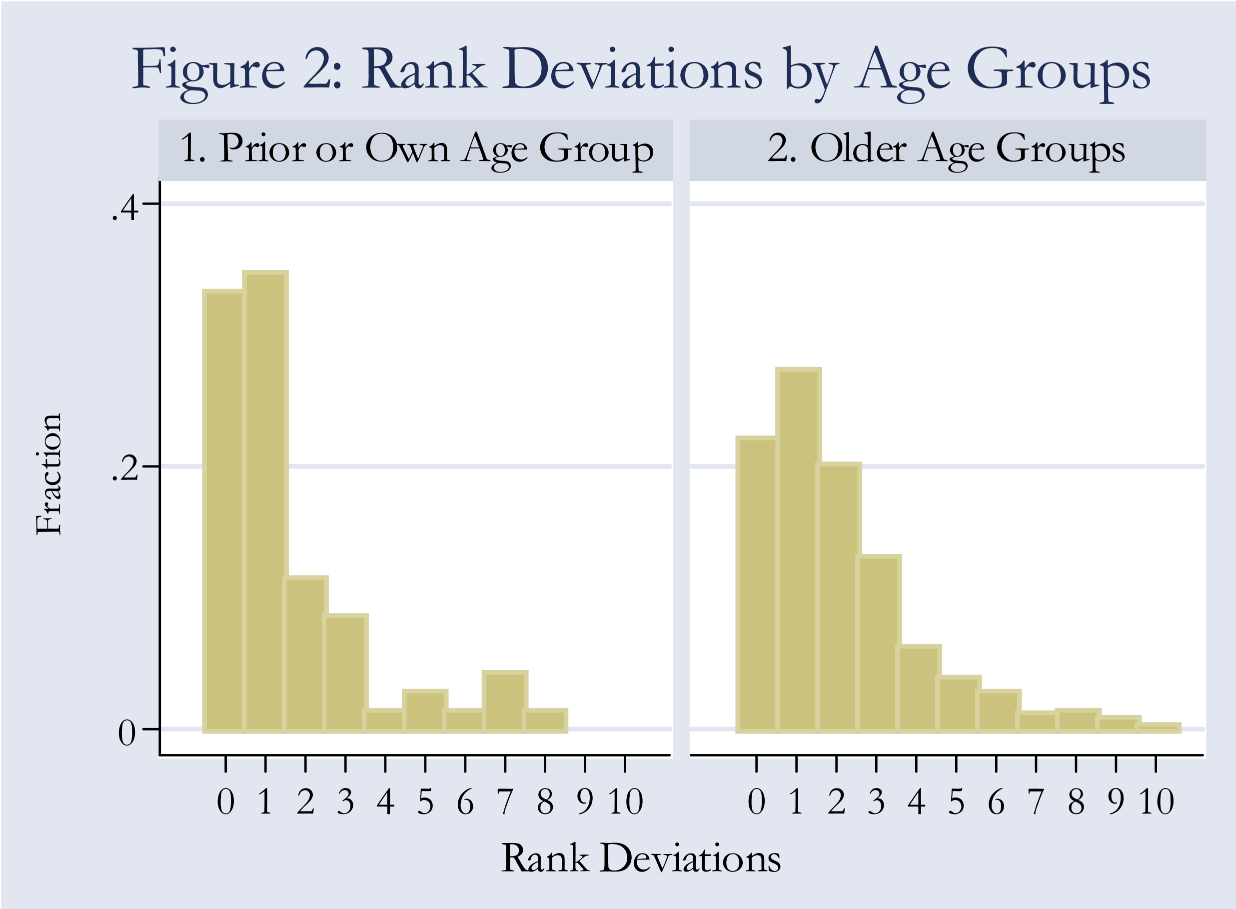
Statistic Type	Mean	Standard Deviation	Minimum	Maximum
Overall	2.007	1.898	0	10
Between		0.383	1.46	3.21
Within		1.859	1.20	10.17

[10] A zero here means that an individual estimated the correct rank for a particular cell. The maximal error that could be made was 11. Hence the values of this variable are bounded between 0 and 11.



The overall error in ranks is 2.01, but ranges between 0 and 10. The overall standard deviation is nearly 2, and most of this is "within individuals." This means that the variation in accuracy occurred much more within the 48 responses of each subject, rather than across subjects. Aggregate, unconditional statistics such as the mean and standard deviation are of only limited use when evaluating these results, since there are many factors at the individual level that should be controlled for in order to understand the observed behavior.

Figure 1 displays the entire distribution of rank errors. The distribution is asymmetric, with many low values and a few extremely high values. This indicates that there are some individuals that are particularly inaccurate and/or some cells that are particularly difficult to rank. Based on the decomposition of standard deviation in Table 2, we conjecture that it is the latter, but our regression



analysis will identify that.

Figure 2 shows the distribution of rank errors when one just looks at the rankings of risks for the own and prior age group of the individual (panel 1) or for all older age groups (panel 2). There is a clear shift up in the distribution for older age groups, implying larger errors. Thus it appears that the two hypothesized age group effects have a substantial joint effect on the accuracy of responses. Again, however, these are results that are only conditioned on one variable, and a complete analysis must be multi-variate and conditioned on characteristics of the individual as well as other characteristics of the task.

Our statistical analysis employs a negative binomial model, since the dependent variable consists of integers between 0 and 10. The negative binomial model is popular for such "count data" since it avoids some of the restrictive assumptions of some popular specifications.[11]

We employ population-averaged estimation methods known as Generalized Estimating Equations (GEE) instead of the customary random-effects specifications for unobserved individual effects. These methods grew out of the General Linearized Models (GLM) tradition initiated by Nelder and Wedderburn [1972], primarily due to an insight from Liang and Zeger [1986]. Excellent reviews of GLM and GEE, stressing their connections, can be found in Hardin and Hilbe [2001][2003], respectively. The relationship between the population-average approach and the random effects (and fixed effects) approach is summarized by Hardin and Hilbe [2003; p.49]: "There are two classifications of models [...] for addressing the panel structure of data. A population-averaged model is one which includes the within-panel dependence by averaging effects over all panels. A subject-specific model is one which addresses the within-panel dependence by introducing specific panel-level random components. A population-averaged model, also known as a marginal model, is obtained through introducing a parameterization for a panel-level covariance. The panel-level covariance (or correlation) is then estimated by averaging across information from all of the panels. A subject-specific model is obtained through the introduction of a panel effect. While this implies a panel-level covariance, each panel effect is estimated using information only from the specific panel. Fixed-effects and random-effects models are subject specific." Population-averaged estimates are relatively popular in epidemiology and some areas of operations research, perhaps due

[11] The Poisson model is the simplest possible count model, and assumes that the mean and variance of the errors are equal. In practice the variance of the errors is often larger than the mean. When the variance is larger than the mean, there are two popular extensions of the Poisson model. One is the over-dispersed Poisson model, in which an extra parameter is included which estimates the scale of the variance relative to the mean. This parameter estimate is then used to correct for the effects of the larger variance on the standard errors of the coefficient estimates. The other alternative is the negative binomial distribution, which is an extension of the Poisson distribution in which the distribution's parameter is itself considered a random variable. The estimated variation of this parameter can account for an error variance of the underlying model that differs from the mean.

to the earlier popularity of GLM in those fields.[12]

Table 3 reports the results of negative binomial panel regressions of the rank errors, controlling for all of the variables referred to earlier as well as a set of individual socio-demographic characteristics defined below. The first three columns of numbers report results using the full sample (of subjects facing salient rewards). The second three columns of numbers report estimates based on restricting the sample to responses that pass the 2% "perceptual difficulty filter" described above.

The variables are defined as follows. The dependent variable, DEV, is the rank error of each individual. MYAGEGRP is a dummy variable picking out if responses fall within the age group of the individual. CUAGEGRP is a dummy variable picking out if the responses fall in an age group that the individual has lived through; it is zero for the age group that the individual is living in now (thus there is no overlap in CUAGEGRP and MYAGEGRP). The age distribution in our sample is skewed towards young adults. However, 64% of the sample were in the age group 15-24, and 33% were in the age group 25-44, so we do have some variation in the sample responses across the two lowest age groups. NUM is a variable that reflects the fraction of deaths from that cause of death in the general adult population, as a measure of the "numerosity" of that cause of death; we express NUM as a percent. AGE is a variable measuring age in years. MALE is a dummy variable picking out males. BLACK is a dummy variable picking out individuals that are either African or African-American, and

[12] The GEE estimates require, amongst other things, that the mean effects and the correlation structure be correctly specified in order for the estimates to be valid. Under the full set of assumptions needed for the random effects specification, GEE estimates are not the most efficient, but they are likely to be more efficient than pooled estimators that ignore the intra-panel correlation (e.g., Wooldridge [2002; p.487]). On the other hand, GEE does not require the zero-correlation assumption of the random effects specification that is the focus of Hausman test and that is often rejected. Random-effects estimation yields virtually identical results, as it happens, but exhibits some numerical instability in convergence. We would not use fixed-effects estimation even if it were feasible, since it would sweep out all demographic characteristics and we are interested in understanding the source of any heterogeneity of response. All estimation was undertaken using version 8.2 of *Stata* (StataCorp [2003a][2003b]).

ASIAN likewise picks out individuals that are either Asian or Asian-American. BUSINESS is a dummy variable picking out individuals that have a major in the Moore School of Business. SENIOR is a dummy variable picking out individuals that are undergraduate seniors. HONORS is a dummy variable identifying individuals that are in the University of South Carolina Honors College.[13] GRAD is a dummy variable indicating if the individual is in some graduate program. BACHELOR is a dummy variable indicating that the individual only aspires to complete a bachelor's degree; PHD is a dummy variable indicating individuals that aspire to complete a Ph.D. degree. CFATHER and CMOTHER are dummy variables identifying individuals whose father or mother, respectively, attended college. AID is a dummy variable identifying individuals that receive some form of financial aid to attend college. USCITIZEN is a dummy variable identifying individuals that are U.S. citizens. SVISA is a dummy variable identifying individuals that are holding a student visa. MARRIED is a dummy variable identifying individuals that are married or who have been married. GPAA is a dummy variable identifying individuals that held, at the time of the experiment, a high GPA average defined as above 3.25.

We find strong support for our hypotheses. In both regressions the variable NUM is statistically significant and has the expected (negative) sign. Since the coefficients of a negative binomial regression show the percentage change in the expected value of the dependent variable for a unit change in continuous explanatory variables,[14] the coefficient on NUM in Table 3 for the full sample indicates that a 1 percentage point increase in the fraction of deaths of a specific cause is associated with a 3.0% drop in the expected subjective ranking error.

[13] The USC Honors College is often identified as a separate group within the University of South Carolina for administrative purposes, and any student that is in it would normally identify it before identifying themselves as a freshman or sophomore or senior.

[14] See Cameron and Trivedi [1998; p.81].

Table 3: Negative Binomial Panel Regressions of Rank Errors

Variable	Full Sample (N=2160 over 45 subjects) Estimate	Standard Error	*p*-value	Sub-Sample of Choices Meeting 2% Perceptual Difficulty Filter (N=675 over 45 subjects) Estimate	Standard Error	*p*-value	Mean
			A. Treatment Controls				
MyAgeGrp	-0.001	0.064	0.993	-0.241	0.116	0.037	
CuAgeGrp	-0.106	0.104	0.312	-0.688	0.197	0.000	
Num	-0.030	0.006	0.000	-0.040	0.007	0.000	
			B. Demographic Controls				
age	0.014	0.009	0.136	0.028	0.016	0.074	24.70
male	-0.004	0.069	0.953	0.029	0.121	0.808	0.51
black	0.109	0.139	0.435	-0.110	0.249	0.659	0.04
asian	0.543	0.251	0.031	0.055	0.438	0.900	0.02
business	-0.001	0.067	0.991	0.009	0.115	0.939	0.49
senior	0.056	0.089	0.529	-0.097	0.152	0.521	0.47
honors	-0.118	0.158	0.457	-0.506	0.281	0.072	0.04
grad	-0.093	0.108	0.391	-0.190	0.185	0.306	0.36
bachelor	-0.186	0.114	0.102	-0.170	0.196	0.386	0.11
PhD	-0.079	0.119	0.506	-0.026	0.203	0.899	0.13
Cfather	0.025	0.075	0.734	0.027	0.131	0.835	0.69
Cmother	0.050	0.074	0.496	0.080	0.129	0.536	0.60
aid	0.122	0.076	0.109	0.197	0.133	0.140	0.73
UScitizen	0.008	0.244	0.974	-0.032	0.425	0.941	0.89
Svisa	0.122	0.253	0.630	0.415	0.439	0.344	0.09
Married	-0.068	0.112	0.541	-0.038	0.194	0.843	0.16
GPAa	0.038	0.063	0.547	0.044	0.109	0.685	0.53
Constant	0.280	0.365	0.443	0.157	0.636	0.805	1.00

Note: See text for definitions of all variables, and specification of the statistical model. A Wald χ^2 statistic with 20 degrees of freedom tests the null hypothesis that all coefficients are equal to zero. It has value 54.2 (52.3) in the Full Sample (Sub-Sample), implying *p*-values of 0.001 (0.001).

As we eliminate the causes that are labeled perceptually difficult, which include many of the low frequency causes, subjects are better able to rank the causes of death for their own age group and the immediately prior age groups than for older age groups. The age-group variables are not statistically significant in the full sample in Table 3, but become weakly statistically significant with the expected (negative) sign as the samples are restricted to the sub-samples that pertain to the perceptually simpler responses. This effect is relatively large. The coefficient on MYAGEGRP is -0.24, and since the explanatory variable is a dummy variable the expected error in subjective rankings is $0.78 = e^{-0.24}$ times larger on average when the dummy variable is zero.[15] Thus, subjects estimate their own age group rankings about 22% (=100% - 78%) *better* than they do other age group rankings. With respect to prior age groups, the coefficient of -0.68 for CUAGEGRP implies that subjects have lower rank errors of about 50% for age groups that they have lived through,[16] as compared to other age groups including the one the individual is living in currently. Thus *our subjects are much better at ranking causes of death both in their own age group, and in age groups they have lived through.*

Although very few of the other socio-demographic variables are individually statistically significant, perhaps due to limited variation in some of them, all of the variables in each regression are jointly significant. This is shown at the bottom of Table 3 in the values for the Wald test, which uses a χ^2 test. Similar tests of the joint statistical significance of all the socio-demographic variables[17] show that they are only statistically significant in the regression with the full sample. Thus we conclude that the accuracy of individual beliefs about mortality risks does *not appear to be affected by their socio-demographic characteristics*, after controlling for age-group and "numerosity" effects.

Since it can be argued that familiarity with some causes may not be a function solely of the

[15] See Cameron and Trivedi [1998; p.82].

[16] Since $e^{-0.68} = 0.51$, implying that the errors are about 49% lower.

[17] That is, excluding the effects of MYAGEGRP, CUAGEGRP and NUM.

numerosity, we also tested a small set of dummies for the specific causes Alzheimer's disease, HIV infection, and Diseases of the heart, based on these being representative of death causes and diseases that are often discussed and presented in the media. We find that subjects have a significantly lower ranking error for Alzheimer's disease than average, and when we eliminate the perceptually difficult tasks (including Alzheimer's disease) we find that subjects have a significantly higher ranking error for HIV infection than for the average disease. Incorporating these dummies does not change the qualitative conclusions, however. It appears that our hypotheses are strongly supported by these data.

We also ran regressions to test if the perceptual bias observed in previous experiments, where subjects were asked to report the *frequency* of risks, is present in the ordering task as well. If it is, we would expect to see causes of death with low true ranks to be "over-ranked," and causes of death with high true ranks to be "under-ranked," leading to a correlation between true and reported ranks that is substantially less than 1. Benjamin et al. [2001] report correlation coefficients of 0.03 based on the Lichtenstein et al. [1978] data, and 0.09 for their own data. For the ordering task we estimate a statistically significant coefficient of 0.56, which is much closer to the slope of 1 which would imply a complete absence of perceptual bias.[18]

3. Hypothetical Bias

A. Changes in Experimental Procedures

The hypothetical versions of the survey instrument replaced the text in the original versions

[18] In this case we estimated a panel negative binomial regression of the elicited ranks on the logarithm of true ranks, controlling for all of the other factors considered in Table 3. Since the explanatory variable here is in logarithms, we can interpret the coefficient directly as an ordinary elasticity (see Cameron and Trivedi [1998; p.81]).

presented earlier which described the salient reward for accuracy. The replacement text was much simpler:

> You will be paid $10 for your time. We would like you to try to rank these as accurately as you can, compared to the official tabulations put out by the U.S. Department of Health. When you have finished please check that all cells in the table below are filled in.

The experiment was otherwise administered identically to those reported earlier.

Subjects were again recruited from classes at the Moore School of Business in early October 2001, a week or so after the previous experiment. One was an advanced Accounting class, and the other was an Economics Principles class. All experiments were administered by the same experimenter (Harrison). These two classes were from very different disciplines, and it was assumed that there was no contact between the subjects regarding the survey questions. Since the payments were not salient, there would indeed be no incentive for the subjects to pass on any information between classes, even if they knew the other students. All of the hypothetical responses were collected after the real responses. There were 27 subjects in the first hypothetical session, and 68 in the second hypothetical session, for a total of 95 subjects. After removing subjects that failed to complete the survey in some respect, there are 91 remaining subjects. The age distribution in this sample is skewed somewhat more towards younger adults than the sample facing real responses: 86% of the sample were in the age group 15-24, and 13% were in the age group 25-44, so we do still have some variation in the sample responses across the two lowest age groups.

B. Experimental Results

Table 4 summarizes the raw responses for the subjects in the hypothetical sessions. The rank errors for the hypothetical (H) sessions are virtually identical to those in the real (R) sessions. The average rank error in the H sessions is 2.15, compared to 2.00 in the R sessions. Moreover, the standard deviation in the H sessions is 1.95, which is also close to the 1.90 for the R sessions. Although there has been some evidence to suggest that average H responses *might* be the same as R responses in *some* settings, it is common to see a significantly higher variance in H responses as noted earlier.

Table 4: Descriptive Statistics for Hypothetical Rank Deviations

Sample size = 4368, consisting of 48 decisions by 91 subjects

Statistic Type	Mean	Standard Deviation	Minimum	Maximum
Overall	2.154	1.950	0	11
Between		0.369	1.29	3.08
Within		1.915	0.93	10.49

Turning to the regression analysis, reported in the Appendix[19], the conclusion from the raw descriptive statistics is confirmed. The H responses result in even stronger confirmation of the hypothesis tests proposed earlier. Specifically, the results for all responses show that the coefficients on the variables MYAGEGRP, CUMAGEGRP and NUM are all negative and statistically significant.

It is also possible to pool the H and R responses into one statistical model and test for the effect of the salience of rewards in the R sessions while controlling for all of the other variables. The dummy variable indicating the real responses is statistically insignificant in both of the samples: the full sample of 48 responses for each subject, and in the sample which filter out the perceptually tough cells. There is no difference between the H and R responses.

[19] Available at the ExLab Digital Archive, http://exlab.bus.ucf.edu.

4. Conclusions

Our results demonstrate that it is possible to design a survey instrument to differentiate the beliefs that subjects have about mortality risks of people in their own age group from those of other age groups. The effect of controlling for the beliefs referring to own vs. other age groups is important. Similarly, the general frequency of the type of death being elicited seems to be important. Both effects are as hypothesized, based on a presumption that individuals have better information about mortality risks that are relevant to them, such as those for their own age group. Apart from the age of the subject, individual demographic characteristics in our sample do not appear to influence the accuracy of responses.

The conclusion from the hypothetical survey variant is a surprise, given the extensive literature on the extent of hypothetical bias: the responses obtained in this hypothetical setting[20] are statistically identical to those found in a real setting. One feature of the vast literature on hypothetical bias is that it deals exclusively with *valuation* tasks, rather than *ranking* tasks. The experimental task studied here is a ranking task. It is possible that the evidence on hypothetical bias in valuation settings does not apply so readily to ranking tasks.[21]

---

[20] The hypothetical setting implemented here should be better referred to as a non-salient experiment. We did reward subjects for participating, with a fixed show-up fee of $10. The hypothetical surveys popular in the field rarely reward subjects for participating, although it has occurred in some cases. Our only point is that there could be a difference between our non-salient experiment and "truly hypothetical" experiments.

[21] One account of hypothetical bias that is consistent with these data runs as follows. Assume that subjects come into an experiment task and initially form some beliefs as to the "range of feasible responses," and that they then use some heuristic to "narrow down" a more precise response within that range. It is plausible that hypothetical bias could affect the first step, but not be so important for the second step. If that were the case, then a task that constrained the range of feasible responses, such as our ranking task that restricts the subjects to choose ranks between 1 and 12, might not suffer from hypothetical bias. On the other hand, a valuation task might plausibly elicit extreme responses in a hypothetical setting, as subjects note that they could just as easily say that they would pay nothing as say that they would pay a million dollars. In this setting there is no natural constraint, such as comparing to one's budget, to constraint feasible responses. Hence the second stage of the posited decision process would be applied to different feasible ranges, and even if the second stage were roughly the same for hypothetical and real tasks, if the first stage were

References

Benjamin, Daniel K., and Dougan, William R., "Individuals' Estimates of the Risks of Death: Part I – A Reassessment of the Previous Evidence," *Journal of Risk and Uncertainty*, 15, 1997, 115-133.

Benjamin, Daniel K.; Dougan, William R.; and Buschena, David, "Individuals' Estimates of the Risks of Death: Part II – New Evidence," *Journal of Risk and Uncertainty*, 22, 2001, 35-57.

Cameron, A. Colin, and Trivedi, Pravin K., *Regression Analysis of Count Data* (New York: Cambridge University Press, 1998).

Coller, Maribeth, and Williams, Melonie B., "Eliciting Individual Discount Rates," *Experimental Economics*, 2, 1999, 107-127.

Croson, Rachel, "Thinking like a game theorist: Factors affecting the frequency of equilibrium play," *Journal of Economic Behavior and Organization*, 41(3), 2000, 299-314.

Cummings, Ronald G.; Elliott, Steven; Harrison, Glenn W., and Murphy, James, "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy*, 105(3), June 1997, 609-621.

Cummings, Ronald G.; Harrison, Glenn W., and Rutström, E. Elisabet, "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible?" *American Economic Review*, 85(1), March 1995, 260-266.

Dufwenberg, Martin and Gneezy, Uri, "Measuring beliefs in an experimental lost wallet game," *Games & Economic Behavior*, 30, 2000, 163-182.

Hardin, James, and Hilbe, Joseph, *Generalized Linear Models and Extensions* (College Station, TX: Stata Corporation, 2001).

Hardin, James, and Hilbe, Joseph, *Generalized Estimating Equations* (Boca Raton: Chapman & Hall/CRC, 2003).

Harrison, Glenn W., "Experimental Evidence on Alternative Environmental Valuation Methods," *Working Paper 04-01*, Department of Economics, College of Business Administration, University of Central Florida, 2004; forthcoming, *Environmental & Resource Economics*.

Harrison, Glenn W., and Rutström, E. Elisabet, "Experimental Evidence of Hypothetical Bias in Value Elicitation Methods," in C.R. Plott and V.L. Smith (eds.), *Handbook of Experimental Economics Results* (Princeton: Princeton University Press, 2005).

---

sufficiently different then the final response could be very different. This is speculation, of course. The experiment considered here does not provide any evidence for this specific thought process, but it does serve to rationalize the results.

Liang, K-Y., and Zeger, S.L., "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 1986, 13-22.

McDaniel, Tanga M., and Rutström, E. Elisabet, "Decision Making Costs and Problem Solving Performance," *Experimental Economics*, 4(2), October 2001, 145-161.

McKelvey, Richard D., and Page, Talbot, "Public and Private Information: An Experimental Study of Information Pooling," *Econometrica*, 58(6), November 1990, 1321-1339.

Nelder, J.A., and Wedderburn, R.W.M., "Generalized Linear Models," *Journal of the Royal Statistical Society – Series A*, 135(3), 1972, 370-384.

Lichtenstein, Sarah; Slovic, Paul; Fischoff, Baruch; Layman, Mark; and Combs, Barbara, "Judged Frequency of Lethal Events," *Journal of Experimental Psychology: Human Learning and Memory*,4(6), 1978, 551-578.

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

Morgan, M. Granger, and Henrion, Max, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis* (New York: Cambridge University Press, 1990).

Neill, Helen R.; Cummings, Ronald G.; Ganderton, Philip T.; Harrison, Glenn W., and McGuckin, Thomas, "Hypothetical Surveys and Real Economic Commitments", *Land Economics*, 70(2), May 1994, 145-154.

Nyarko, Yaw, and Schotter, Andrew, "An Experimental Study of Belief Learning Using Elicited Beliefs," *Econometrica*, 70(3), May 2002, 971-1005.

Savage, Leonard J., "Elicitation of Personal Probabilities and Expectations," *Journal of American Statistical Association*, 66, December 1971, 783-801.

StataCorp, *Stata Statistical Software: Release 8.0* (College Station, TX: Stata Corporation, 2003a).

StataCorp, *Stata Cross-Sectional Time-Series Reference Manual: Release 8.0* (College Station, TX: Stata Corporation, 2003b).

U.S. Department of Health & Human Services, *Health United States 1986-96* Washington, D.C.: U.S. Government Printing Office, DHHS Publication # (PHS) 93-1232 (Annual), 1997.

U.S. Environmental Protection Agency (USEPA), *The Benefits and Costs of the Clean Air Act, 1970 to 1990*. United States Environmental Protection Agency, Office of Policy Analysis and Review, Document # A.97.20, October 1997.

Viscusi, W. Kip, "Do smokers underestimate risks?" *Journal of Political Economy*, 98(6), 1990, 1253-1269.

Viscusi, W. Kip, "The Value of Risks to Life and Health," *Journal of Economic Literature*, 31(4), December 1993, 1912-1946.

von Winterfeldt, Detlof, and Edwards, Ward, *Decision Analysis and Behavioral Research* (New York: Cambridge University Press, 1986).

Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data* (Cambridge, MA: MIT Press, 2002).

Appendix : Additional Materials (NOT FOR PUBLICATION)

A.1 Survey Questions Eliciting Individual Characteristics

SOME QUESTIONS ABOUT YOU

What is your age? ___________

What is your sex? FEMALE MALE

What is your race? Please pick one:

- ❑ White
- ❑ African-American
- ❑ African
- ❑ Asian-American
- ❑ Asian
- ❑ Hispanic-American
- ❑ Hispanic
- ❑ Mixed race
- ❑ Other

Which category best describes your current major? Please pick one:

- ❑ Economics
- ❑ Business and Management, other than Economics
- ❑ Education
- ❑ Engineering
- ❑ Health Professions
- ❑ Public Affairs or Social Services
- ❑ Biological Sciences
- ❑ Math, Computer Sciences, or Physical Sciences
- ❑ Social Sciences or History
- ❑ Humanities
- ❑ Psychology
- ❑ Other Fields

What is your student status? Please pick one:

- ❑ Freshman
- ❑ Sophomore
- ❑ Junior
- ❑ Senior
- ❑ Honors
- ❑ Masters
- ❑ Doctoral
- ❑ Non-student

What is the **highest** level of education you expect to **complete**? Please pick one:

❑ Bachelor's degree
❑ Master's degree
❑ Doctoral degree
❑ First professional degree

What was the highest level of education that your **father** (or male guardian) **completed**? Please pick one:

❑ Less than high school
❑ GED or High School Equivalency
❑ High school
❑ Vocational or trade school
❑ College or university

What was the highest level of education that your **mother** (or female guardian) **completed**? Please pick one:

❑ Less than high school
❑ GED or High School Equivalency
❑ High School
❑ Vocational or trade school
❑ College or university

In financing your current degree, have you received any financial aid from grants, scholarships or loans to help defray the costs?

❑ Yes
❑ No

What is your citizenship status in the United States?

❑ U.S. Citizen
❑ Resident Alien
❑ Non-Resident Alien
❑ Other Status

Are you a foreign student on a Student Visa? YES NO

Are you currently...

❑ Single and never married?
❑ Married?
❑ Separated, divorced or widowed?

On a 4-point scale, what is your current GPA if you are doing a Bachelor's degree, or what was it when you did a Bachelor's degree? This GPA should refer to all of your coursework, not just the current year. Please pick one:

- ❏ Between 3.75 and 4.0 GPA (mostly A's)
- ❏ Between 3.25 and 3.74 GPA (about half A's and half B's)
- ❏ Between 2.75 and 3.24 GPA (mostly B's)
- ❏ Between 2.25 and 2.74 GPA (about half B's and half C's)
- ❏ Between 1.75 and 2.24 GPA (mostly C's)
- ❏ Between 1.25 and 1.74 GPA (about half C's and half D's)
- ❏ Less than 1.25 (mostly D's or below)
- ❏ Have not taken courses for which grades are given.

A.2 Regression Analyses for the Hypothetical Experiments

Table 5: Summary Statistics of Subject Pool

A. Hypothetical Responses

Variable	Obs	Mean	Std. Dev.	Min	Max
age	91	21.46154	4.578471	18	49
male	91	.3846154	.4891996	0	1
black	91	.0769231	.2679457	0	1
asian	91	.1208791	.3277928	0	1
business	91	.6813187	.4685467	0	1
senior	91	.021978	.147424	0	1
honors	91	0	0	0	0
grad	91	.2967033	.4593354	0	1
bachelor	91	.1428571	.3518658	0	1
PhD	91	.1208791	.3277928	0	1
Cfather	91	.6593407	.4765566	0	1
Cmother	91	.6373626	.4834249	0	1
aid	91	.8571429	.3518658	0	1
UScitizen	91	.8901099	.3144855	0	1
Svisa	91	.0659341	.2495417	0	1
Married	91	.1098901	.3144855	0	1
GPAa	91	.7362637	.4430993	0	1

B. Real Responses

Variable	Obs	Mean	Std. Dev.	Min	Max
age	45	24.66667	5.049752	19	47
male	45	.5111111	.505525	0	1
black	45	.0444444	.2084091	0	1
asian	45	.0222222	.1490712	0	1
business	45	.4888889	.505525	0	1
senior	45	.4666667	.504525	0	1
honors	45	.0444444	.2084091	0	1
grad	45	.3555556	.4840903	0	1
bachelor	45	.1111111	.3178209	0	1
PhD	45	.1333333	.3437758	0	1
Cfather	45	.6888889	.4681794	0	1
Cmother	45	.6	.4954337	0	1
aid	45	.7333333	.4472136	0	1
UScitizen	45	.8888889	.3178209	0	1
Svisa	45	.0888889	.287799	0	1
Married	45	.1555556	.3665289	0	1
GPAa	45	.5333333	.504525	0	1

Table 6: Negative Binomial Panel Regression of Rank Errors with Full Sample (Hypothetical and Real Responses Included)

```
GEE population-averaged model                   Number of obs      =      6528
Group variable:                         id      Number of groups   =       136
Link:                                  log      Obs per group: min =        48
Family:           negative binomial(k=1)                       avg =      48.0
Correlation:                 exchangeable                      max =        48
                                                Wald chi2(22)      =     66.77
Scale parameter:                         1      Prob > chi2        =    0.0000
```

Dev	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
real	-.0072146	.0766367	-0.09	0.925	-.1574198	.1429906
Session	.0751566	.1100971	0.68	0.495	-.1406297	.2909428
MyAgeGrp	-.0869828	.0359671	-2.42	0.016	-.157477	-.0164885
CuAgeGrp	-.1764609	.0757472	-2.33	0.020	-.3249226	-.0279992
Num	-.0179722	.0030572	-5.88	0.000	-.0239642	-.0119803
age	.0065256	.0062785	1.04	0.299	-.00578	.0188312
male	.018472	.0385861	0.48	0.632	-.0571553	.0940994
black	.0127661	.0731562	0.17	0.861	-.1306175	.1561496
asian	.1986004	.0847896	2.34	0.019	.0324158	.364785
business	-.0169729	.0402799	-0.42	0.673	-.09592	.0619743
senior	.0076509	.0832243	0.09	0.927	-.1554657	.1707675
honors	-.1268314	.1734053	-0.73	0.465	-.4666995	.2130367
grad	-.0731991	.1005164	-0.73	0.466	-.2702077	.1238095
bachelor	-.0514386	.054307	-0.95	0.344	-.1578783	.0550011
PhD	.0412503	.0582833	0.71	0.479	-.0729829	.1554834
Cfather	-.0228848	.0464355	-0.49	0.622	-.1138967	.0681271
Cmother	-.0137892	.0441697	-0.31	0.755	-.1003603	.0727819
aid	.0913061	.0516874	1.77	0.077	-.0099993	.1926114
UScitizen	.0998461	.1015864	0.98	0.326	-.0992596	.2989518
Svisa	.1692657	.1236228	1.37	0.171	-.0730306	.411562
Married	-.0445115	.0723089	-0.62	0.538	-.1862344	.0972114
GPAa	-.0169961	.0427258	-0.40	0.691	-.1007371	.0667449
_cons	.4999929	.2140993	2.34	0.020	.0803659	.9196199

Legend: variable "real" equals 1 if the response is real, 0 otherwise.

Table 7: Negative Binomial Panel Regression of Rank Errors with Sub-Sample Meeting 2% Perceptual Difficulty Filter (Hypothetical and Real Responses Included)

```
GEE population-averaged model                   Number of obs      =      2040
Group variable:                         id      Number of groups   =       136
Link:                                  log      Obs per group: min =        15
Family:             negative binomial(k=1)                     avg =      15.0
Correlation:                  exchangeable                     max =        15
                                                Wald chi2(22)      =     97.71
Scale parameter:                         1      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         Dev |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        real |   .0500815   .1166226     0.43   0.668    -.1784947    .2786577
     Session |   .0440798   .1667069     0.26   0.791    -.2826596    .3708192
    MyAgeGrp |  -.3667346   .0659152    -5.56   0.000     -.495926   -.2375432
    CuAgeGrp |  -.7132138   .1399967    -5.09   0.000    -.9876022   -.4388254
         Num |  -.0252887   .0035419    -7.14   0.000    -.0322307   -.0183467
         age |   .0182894   .0096714     1.89   0.059    -.0006662    .0372451
        male |   .0606301   .0587641     1.03   0.302    -.0545454    .1758055
       black |  -.0595255   .1119529    -0.53   0.595    -.2789492    .1598982
       asian |   .1218682   .1288401     0.95   0.344    -.1306539    .3743902
    business |    .021129   .0613906     0.34   0.731    -.0991943    .1414523
      senior |  -.1470708   .1263131    -1.16   0.244      -.39464    .1004983
      honors |  -.5222536    .272666    -1.92   0.055    -1.056669    .0121619
        grad |   -.135989   .1517112    -0.90   0.370    -.4333375    .1613596
    bachelor |  -.0763473   .0829035    -0.92   0.357    -.2388351    .0861405
         PhD |   .0866087   .0881933     0.98   0.326     -.086247    .2594643
     Cfather |  -.1323138   .0705889    -1.87   0.061    -.2706654    .0060378
     Cmother |   .0319186   .0672982     0.47   0.635    -.0999835    .1638207
         aid |     .17531   .0792528     2.21   0.027     .0199773    .3306427
   UScitizen |   .0170728   .1536546     0.11   0.912    -.2840847    .3182303
       Svisa |   .2559937   .1872836     1.37   0.172    -.1110755    .6230628
     Married |   .0089065   .1098235     0.08   0.935    -.2063436    .2241566
        GPAa |  -.0330905    .064994    -0.51   0.611    -.1604763    .0942954
       _cons |   .4539593    .327242     1.39   0.165    -.1874232    1.095342
------------------------------------------------------------------------------
```

A.3 Data Analysis and *Stata* Code

All files pertaining to the experiments are available in machine-readable format at project "Eliciting Mortality Risk Orderings" at the ExLab Digital Archive located at http://exlab.bus.ucf.edu. For convenience, each file is available in one overall ZIP archive called CAUSES OF DEATH – PUBLIC VERSION.ZIP.

The survey instruments are available in *WordPerfect* and *Adobe PDF* format. The raw data was entered from the hard-copy survey responses into a *Quattro Pro* spreadsheet DEATH.WB3, which also contains the calculations embodied in Table 1. The program *DBMS/COPY* was then used to convert the responses to *Stata* data file DEATH.DTA, provided on the web page. The actual names of the subjects were needed for payment and university reimbursement purposes, but are kept privately to ensure confidentiality. An alternative version, called NAMESANON.DTA, is provided on the web page and masks the true name. This allows the computer programs to run with some dummy names in place.

The main statistical analysis is undertaken by two *Stata* programs, called CODDATA.DO and CODEVAL.DO. Version 8.2 of *Stata* was employed. The first program collates all of the data and calculates profits for each subject; it was used to determine payments to each subject. The second program undertakes the statistical analyses reported here.