AN EMPIRICAL MODEL OF PERCEIVED MORTALITY RISKS FOR SELECTED UNITED STATES ARSENIC HOT SPOTS*

[Forthcoming Risk Analysis, Accepted May 2010]

To N. Nguyen, School of Engineering, University of Guelph
Paul M. Jakus, Dept. of Applied Economics, Utah State University
Mary Riddel, Dept. of Economics, University of Nevada, Las Vegas
W. Douglass Shaw, Dept. of Agricultural Economics, Texas A&M University

Abstract: Researchers have long recognized that subjective perceptions of risk are better predictors of choices over risky outcomes than science-based or expert’s assessment of risk. More recent work suggests that uncertainty about risks also plays a role in predicting choices and behavior. In this manuscript we develop and estimate a formal model for an individual’s perceived health risks associated with arsenic contamination of their drinking water. The modeling approach treats risk as a random variable, with an estimable probability distribution whose variance reflects uncertainty. The model we estimate uses data collected from a survey given to a sample of people living in arsenic prone areas in the United States. The findings from this paper support the fact that scientific information is essential to explaining the mortality rate perceived by the individuals, but uncertainty about the probability remains significant.

Keywords: Arsenic, Mortality rate, Perceived risk, Ambiguity.

* Shaw is contact author (wdshaw@tamu.edu). We knowledge funding for data collection and researcher support from the U.S. Environmental Protection Agency, STAR grant program (#RJ8RJ7), and Jakus and Shaw receive additional support via the U.S.D.A. W-2133 Research project. The authors wish to thank T.A. Cameron, J.R. DeShazo, P. Jindapon, K.Moeltner, L.Schauer and the survey team at PA Consulting, V. K. Smith, K. Stoddard, and RT.Woodward for comments and help on the manuscript, the survey methodology, and data collection. We also thank two reviewers and two editors of this journal, Greg Poe, Shelby Gerking, Glenn Harrison and other participants at seminars at the University of Alabama, Central Florida, Oregon, Texas A&M, Utah State University, Vassar College, and participants in the 2008 W-2133 conference for their comments on this, or closely related work. The usual caveats apply.
1. Introduction

In this paper we develop and estimate an empirical model of subjective mortality risks that assumes individuals may be uncertain about risks associated with arsenic contamination of drinking water. Eliciting and measuring subjective risks is important because risk researchers long have recognized that subjective or individually-perceived risks may be much more important in determining behavior than science-based calculations of risks. More recent research has shown that both risk and uncertainty affect behavior and willingness to pay to reduce risk. For example, Riddell and Shaw (2006) and Cameron (2005) found that risk and uncertainty affect the willingness to accept risks from nuclear-waste transport and climate change, respectively.

Despite the potential importance of risk and uncertainty in risky choice models, the majority of the subjective risk literature assumes that people can form point estimates of risk, with no uncertainty (or ambiguity) assigned to the risk estimate [see the discussion of modeling risks by Wilcox, 2008; or the extensive discussion of how subjective risks are elicited in the experimental laboratory by Harrison and Rutström, 2008]. But uncertainty about subjective risks and, indeed, objectively measured risks, is appropriate in many settings, especially if the environmental, ecological, or epidemiological context of the risky event means that risks are inherently uncertain. Exposure to relatively low concentrations of arsenic in drinking water is one such setting, and is the focus of this study.
Arsenic is a toxin, with long-term consumption of arsenic-contaminated water at concentration levels above 50 parts per billion (ppb) known to cause skin damage, problems with circulatory systems, and lung or bladder cancer. In response to growing evidence that even small amounts of arsenic increase mortality and morbidity, the regulatory standard was recently lowered to 10 ppb (US EPA 2006) from the previous standard of 50 ppb. Nevertheless, the exact dose-mortality relationship remains uncertain, especially for concentrations between 10 and 50 ppb. Some analysts (e.g., Burnett and Hahn 2001; Wilson 2001) have raised concerns about the data and the methodology used by the EPA to estimate the risks of low-level exposure, and doubt is often cast on inferences for human effects based on animal and epidemiological studies. Other critics believe that the dose-response relationship from arsenic should be nonlinear rather than the linear model used by EPA. Estimation of the dose-response relationship is further complicated by factors such as individual consumption of contaminated water and related choices, such as cigarette smoking, that can exacerbate the effects of arsenic-contaminated water. For example, the National Research Council [NRC] (2001) has estimated that smokers have at least two times the mortality risk of non-smokers because both arsenic and smoking increase the risk of lung cancer. Such confounding factors make it difficult for scientists to offer exact risk estimates for a given exposure to arsenic; we would therefore expect that perceived risks held by non-scientists should also reflect this uncertainty.

Gilboa, Postelwaite and Schmeidler (2008) have recently examined the role of uncertainty in risk analysis in the economics arena, and our study provides a response to their recommendation for more empirical modeling techniques that address uncertain, as opposed
to known, risks. This is continuing the trend in risk analysis began long ago in other fields (see Morgan and Henrion 1990; Casman et al. 1999). We present an empirical analysis of perceived risks of consuming drinking water in regions of the United States that do not meet the new EPA arsenic standard. Risk perceptions [Fischhoff (2009) notes that psychologists often distinguish between risk perceptions and risk judgments, so we advise the reader that we use these interchangeably here] are directly elicited using a telephone-mail-telephone survey format that permitted both written and verbal communication of exposure levels and objective risks to respondents. Survey responses are analyzed using a straightforward statistical model based on the concept that an individual uses information to form some estimate of perceived risk but may still have some degree of uncertainty. The “induced distribution” approach (Heckman and Willis, 1977) leads to models of a point estimate of perceived risk and a measure of the variance (uncertainty or ambiguity). Each model is parameterized using variables believed to influence either the central tendency or the variance. Both the central tendency and variance models allow for heterogeneity in risk beliefs and the distribution of the perceived risk across individuals (e.g., Dosman et al. 2001).

In what follows, Section 2 summarizes the methods used to collect the data and elicit risks from the sample of respondents. In Section 3, the empirical model of perceived risks is presented. The results of the model are discussed in Section 4, and the last section offers some conclusions, and suggestions for extensions of the model.

2.0 The sample and survey of perceived risks
A more complete description of the survey methodology used to elicit perceived risks may be found in Nguyen (2008) and more recently in Jakus et al. (2009), but we briefly describe the key features here. Our sample consists of households living in four communities exposed to arsenic levels in excess of the new EPA standard of 10 ppb at the time of the study (late 2006). Table 1 provides information on drinking water sources and arsenic exposure, including the mean and range of contamination for the four communities. The public water supply systems that provide water to residents of Albuquerque, Fernley (Nevada) and Oklahoma City were not in compliance with federal standard for arsenic. The Outagamie County/Appleton Wisconsin region was selected for the study because arsenic levels in privately owned wells are known to exceed the federal standard of 10 ppb but are not regulated under federal statutes such as the Safe Drinking Water Act.

A series of focus groups was used to assist in the design the survey instruments and an informational booklet. These were conducted in various locations in the United States, including some that were not eventually chosen as locations for the sampling, though most were in areas where arsenic was a problem (the exception was a focus group held near Las Vegas, Nevada, where there is currently no known risk from exposure to arsenic in drinking water). During these sessions subjects were exposed to several different text and visual formats for communicating arsenic risks. The subjects were video-taped and the tapes studied to detect confusion and emotions involved with particular formats, as both may be important in processing the risk information (see Morgan and Henrion 1990; Morgan et al. 2001). We did not explicitly use a mental model approach, as others have done with hazardous substances (e.g. see the Bostrom et al. 1992 application to radon), but the focus group
sessions helped a great deal in forming the information brochure so that respondents could learn about how arsenic exposure occurs, and the key factors that influence it.

Quite some time ago, risk specialists found that graphical/visual communication tools helped people under risks (see Ibrekk and Morgan 1987), but presentation formats are known to influence risk perceptions, (see Connelly and Knuth 1998), with particular attention paid to the presentation of probabilities that are lower than 0.05, as these small probabilities often have been documented to cause problems for respondents. Our pre-survey focus group explored several words relating to death risks, and various formats for presentation. This work led us to conclude that science-based mortality risks were best understood using a risk ladder, where deaths per 100,000 are presented. Risk ladders depicting death rates have been used for many years as a good risk communication device to enhance peoples’ understanding of mortality risks (see Loomis and DuVair 1993; the later comparison study of risk presentation formats by Connelly and Knuth 1998). Although risk ladders have limitations and are not a panacea for all potential risk communication problems (e.g., Sandman, Weinstein and Miller 1994), other risk communication devices, such as risk grids, also have some problems. Our focus group participants had strong preferences for the ladder over the risk grid, so we decided to focus the risk communication using a ladder, with final content revised to reflect the initial confusion of the focus group subjects about our first version of the risk ladder.

A first-round random-digit dial survey was then used to recruit participants into later survey activities focusing on arsenic exposure and perceived risks. The short first-round telephone survey collected information on the respondent’s concerns about general
environmental risks from atmospheric and water pollutants, how tap water was used in the household, and demographic characteristics. At the conclusion of the first phone call, the respondent was asked if he or she would be willing to participate in a further survey on the issue of arsenic contamination and drinking water. Those who agreed to participate were then mailed a multi-page information booklet that explained the risks of arsenic exposure and how these risks could be mitigated. The booklet also reported arsenic concentrations in the community in which the respondent lived (i.e., the data reported in Table 1) and requested that respondents consider their risk of health problems associated with their consumption of water with arsenic concentrations that violate the new EPA standard. Sufficient time was allowed to elapse for the participant to consider the information contained in the booklet, after which the perceived risk was directly elicited during the final follow-up telephone call.

2.1 Elicitation of risk and uncertainty

The information booklet mailed to each respondent described the sources of arsenic contamination, the effects of long-term exposure, the new 10 ppb EPA standard for arsenic, and the level of arsenic in the drinking water in their community. The booklet then presented detailed information regarding the specific mortality risks of arsenic exposure and the confounding factors that affect an individual’s risk such as daily water consumption, smoking, and exposure to second-hand smoke. Other factors influencing risk included the use of an appropriate filtration system and one’s current health status.

The risk ladder used in the survey brochure is presented in Jakus et al. (2009). The most important rungs on the risk ladder corresponded to (1) baseline lung cancer risks (60 deaths per 100,000 people), (2) lung cancer risks associated with twenty years of water
consumption at 50 ppb (1000 per 100,000), (3) lung cancer risks of a smoker after twenty
years of water consumption at 50 ppb (2000 per 100,000). The remaining rungs on the ladder
were associated with a series of low risk events (e.g., people are struck by lightning about 5
in a million times) and high risk ones (e.g., automobile accidents kill about 280 out of
100,000 people) unrelated to arsenic exposure. Alongside each rung on the ladder was a tick
grid corresponding to the ladder rungs; each respondent was asked to put a single mark at the
appropriate mortality risk if they were certain about the risk they faced or, if uncertain, place
two marks on the tick grid associated with the lowest and highest mortality risk they thought
might pertain. The exact statements in the brochure, and those that correspond to these risk
elicitations in the final telephone survey, are included in the appendix to this paper.

2.2 The sample and basic statistics

A total of 748 people completed the first round “recruiter” survey, with 353
respondents completing the second round survey focused on arsenic contamination. Of
those, 201 respondents provided a point estimate of risk, while 96 provided a range of risk,
and 56 ultimately could not (or refused) to provide any estimate of their mortality risk. Table
2 reports descriptive statistics for these three groups (those providing point estimates, those
giving range estimates, and those unable to provide a risk estimate at all). The only
significant difference is that the group unable to state a risk estimate has a lower percentage
of males (45%) relative to the group that made a point estimate of risk (about 61%). This
may be an indication that males are more willing to offer risk estimates than females, or
some might say, that females are more careful and less willing to make such subjective
estimates.
Our final sample excludes subjects that could not offer risk estimates, thus our sample includes 297 subjects, 68% who offered point estimates and 32% who offered ranges. Relative to the first-round recruiter sample, sample selection tests indicate that the final sample is composed of a greater percentage of males and people who were more interested in environmental issues (formal selection models are available upon request). Examination of the data reveals that respondents do have difficulty in assessing all the ways in which one is exposed to arsenic in drinking water. Approximately 65% of respondents reported drinking at least some water from the tap, suggesting that 35% drink no water from their tap, but fully 85% said they used tap water to make beverages such as juice or coffee. This lack of careful consideration in assessing exposure has been documented before (see Shaw, Walker and Benson 2005, for example). About 52% of households in the final sample also report that they treat their drinking water in some way, though the treatment method they use may or may not effectively remove arsenic from their drinking water.

A person reporting a range of risks was assumed to signal some degree of uncertainty about the risks posed by arsenic in drinking water. Figure 1 shows a histogram indicating the mortality risks as reported by respondents, where the midpoint of the risk range is used for respondents who were ambiguous about the risks. The risk distribution is weighted heavily toward relatively low risk levels, with the median of the distribution at about 175 deaths per 100,000 people. This median risk estimate lies between the baseline risk of 60 deaths from bladder and lung cancer with no arsenic exposure (based on current estimates by the Center for Disease Control at their website) and the risk associated with exposure to an arsenic
concentration of 50 ppb (about 1000 deaths); the median estimate is consistent with the arsenic exposure levels in the sampled communities, all of which lie below 50 ppb.

The risk literature is full of examples that help guide us about what factors might influence the median or point estimate of risk. Some factors affect what people judge risks to be, while others may affect how people arrive at these judgments, i.e. their cognitive skills (see the recent discussion in Fischhoff 2009: he provides age as an example of a factor perhaps determining what people know, but not their cognitive ability). To begin, the median perceived risk likely varies with objective variables that determine the arsenic dose-response relationship. Objective variables include the actual concentration of arsenic in the respondent’s drinking water, health and smoking status of the subject, and whether or not the subject works in an occupation that has been shown to increase arsenic-related mortality risks. We expect the actual concentration of arsenic in the subject’s drinking water to be positively correlated with the perceived risk while those subjects who treat their water may perceive lower risks. The subject’s health status likely affects his or her risk perceptions, since those in failing health are more vulnerable to environmental toxins than other, healthier, individuals. Finally, as conveyed to the respondents, smokers and those working in risky occupations are more likely to contract cancer when their water contains arsenic. If subjects accurately incorporate this information, then current and former smokers, as well as those employed in risky occupations, should report higher perceived risk, all other factors being equal.

While the variables outlined above may affect perceptions through the dose-response relationship, median perceived risk estimates have also been shown to vary with
psychosocial factors. For example, gender has frequently been shown to affect perceptions (see Gustafson 1998) or risk judgments, with women typically perceiving higher risk than men, on average (Riddle 2009, Jianakoplos and Bernasek 1998, Slovic 2001). Age may also be a factor: Slovic (2001) and Riddle and Shaw (2006) found negative correlations between age and perceived risk for smoking and nuclear-waste transport risk, respectively. One might believe that cognitive ability of individuals plays an important role in risk perceptions, so that education may also be a determinant of the median perceived risk, though the direction of influence is an empirical matter.

Unlike the median risk, much less is known about what factors influence the perceived variability because studies that account for uncertainty are rare and economic theory provides little guidance. Presumably, objective uncertainty about actual arsenic concentrations in a given water supply may translate into subjective uncertainty about health risks, so we estimate models that include the objective range of arsenic contaminations reported in Table 2. Also, a recent study addressing perceived nuclear-waste transport risk found that psychosocial factors were not important determinants of uncertainty; rather, the number of sources of information the subject used to form risk perceptions and the range of the risks stated by the subject were the sole determinants of the variance of the risk distribution (Riddle 2009). In the current study, we do not have a variable that explicitly measures information held by the subject. However, we do know that arsenic contamination received significant media attention in the Albuquerque media just prior to the survey, so we include indicator variables representing the different survey locales to test for variation in.
uncertainty based on the locale. We also consider whether psychosocial variables (such as age and gender) affect uncertainty of perceived risk.

Table 2 presents the basic statistics for the variables used in estimating the empirical model of risk. The statistical model allows these variables to influence the median perceived risk, the uncertainty a person has about risk, or both. Some 57% of all sample respondents were *Male* and the average *Age* in the sample was 51 years old. The *Education* survey question is categorized into 7 levels, with no high school attendance as the lowest educational level and the category involving the receipt of an advanced university degree as the highest level. The majority of the sample (67%) attended a post-secondary educational institution, with few failing to receive at least a high-school degree.

Respondents were asked to rate their own *Health Status* from excellent to poor (the variable has been coded such that a higher number indicates poorer health.) Over 64% of the sample rated their current health condition as being very good or excellent. Some types of jobs have been shown to increase mortality risks from arsenic exposure because of occupational exposure to other toxins. The survey thus asked the respondents if they were currently employed or had been employed in occupations that scientists believe may increase baseline lung and bladder cancer risks, such as manufacturing paint, textiles, leather, dyes, rubber products or other chemicals, working as a beautician or hairstylist, or working in the printing or aluminum industries. The effects of such occupational exposures were outlined in the mailed information brochure. As shown in Table 2, some 26% of respondents worked or still work in a *Risky Occupation*. 
Questions about smoking behavior were taken verbatim from the widely used national Health and Retirement Survey (HRS). About 46% of the sample reported being a Former Smoker or Current Smoker. Former smokers were defined as anyone who had smoked 100 cigarettes or more in the past, but who did not currently smoke. Former Smokers comprised about 33% of the sample.

3.0 Modeling perceived risks

Any probability distribution with support bounded by zero and one is a candidate for use in modeling the variation in an individual’s stated or elicited perceived risk. Here we use a probability density function (PDF) based on the probit function introduced by first by Heckman and Willis (1977) then extended by Lillard and Willis (2001). The probit function approach is derived as follows. First, denote individual i’s probabilistic belief about his or her own mortality risk by $p_i$. Consider an index function, $I_i$:

$$I_i = m_i + u_i - \epsilon_i \text{ where: } \epsilon_i \sim N(0,1), \ u_i \sim N(0, \sigma_i^2), \ \epsilon_i \perp u_i. \quad [1]$$

Here, $m_i$ represents all of the information used to form the person’s best estimate about the probability, and $u_i$ represents uncertainty about the probabilistic risk. The error term $\epsilon_i$ is assumed to be a standard normal random variable which accounts for measurement error on the part of the researcher. The standard deviation of $u_i$, denoted by $\sigma_i$, represents a summary of the information determining a person’s ambiguity about the risk, where ambiguity might relate, for example, to a lack of information about a risk or uncertainty with respect to how confounding factors affect risk. When an individual has no uncertainty about the risk, then
she is precise, and researchers in turn measure the risk up to the degree associated with the usual measurement error term.

The cumulative distribution and density functions, \( F(p_i) \) and \( f(p_i) \), respectively, are derived from [1] as follows. Let:

\[
p_i = \text{Prob}(I > 0) = \text{Prob}(m_i + u_i - \epsilon_i > 0) = \Phi(m_i + u_i).
\]

[2]

where \( \Phi(\cdot) \) represents the cumulative distribution function (CDF) for a standard normal random variable. The distribution function \( F(p_i) \) can be derived directly from [2]:

\[
F(p_i^*) = \text{Prob}(p_i < p_i^*) = \text{Prob}(\Phi(m_i + u_i) < p_i^*) = \Phi\left(\frac{\Phi^{-1}(p_i^*) - m_i}{\sigma_i}\right)
\]

[3]

Note that when \( \sigma_i = 0 \) there is certainty, so equation [3] does not pertain.

Lillard and Willis (2001) and Riddel (2009) discuss the properties of the distribution in [3] in great detail, but we note a couple points that are salient to the current analysis. The median probability of mortality is equal to \( \Phi(m_i) \). For example, if \( m_i = -2.326 \), the median perceived accident risk is 0.01 implying a death rate of 1,000 per 100,000. Successively smaller (more negative) values of \( m_i \) indicate lower perceived risks, all else equal. Note that for an individual who is certain about the risk \( \sigma_i = 0 \) and \( f(p_i) \) is degenerate at \( \Phi(m_i) \), the median risk.

The expected perceived mortality risk is:
Note that the expected (average) risk is a function of both $\mu$ and $\sigma$ whereas the median is determined solely by $\mu$. The density function $f(p_i)$ will typically be asymmetric for $\sigma > 0$, so that for $\mu < 0$ an increase in $\sigma$ means an increase in the average, $\bar{p}$, as the distribution becomes increasingly right skew. Thus, for subjects expressing uncertainty, the average risk is greater than the median risk.

Our empirical modeling relies upon a likelihood function based on equation [3]. Consider respondent $i$, who marked two rungs on the risk ladder corresponding to the probabilities $p_{i1}$ (lower rung) and $p_{i2}$ (upper rung). Our task is to estimate the model parameters to maximize the likelihood that the probability, $p_i$, falls within the stated range.

To begin, the median of perceived risk, $m_i$, and the ambiguity, as measured by the standard deviation, $\sigma_i$, are parameterized to allow for heterogeneity in the distribution. Specifically, the median and the standard deviation are given by:

$$m_i = X_i \alpha$$

$$\ln \sigma_i = Z_i \beta$$

where $X_i$ and $Z_i$ are vectors of variables that influence the individual’s subjective assessment about the median and variance of the risk, respectively.

The vectors $X_i$ and $Z_i$ may share some common variables; $\alpha$ and $\beta$ reflect weights that the individual put on factors in $X$ and $Z$. Substituting [5] and [6] into equation [3] yields the following:

$$\bar{p} = \int_0^1 \phi \left( \frac{\Phi^{-1}(\pi) - \mu}{\sigma} \right) \phi(\Phi^{-1}(\pi) | \sigma) d\pi$$

[4]
cumulative distribution of the probabilistic belief about the mortality with terms for the explanatory variables in the vectors $X_i$ and $Z_i$:

$$F(p_i) = \Phi \left[ \frac{\Phi^{-1}(p_i) - X_i\alpha}{\exp(Z_i\beta)} \right] \tag{7}$$

The individual contribution of the observation $i$ in the sample likelihood is the likelihood for $p_i$ falling within the stated range, $p \in [p_{i1}, p_{i2}]$, specified as:

$$prob(p_{i1} \leq p_i \leq p_{i2}) = F(p_{i2}) - F(p_{i1}) \tag{8}$$

For those respondents who only marked a single rung, we approximate the likelihood function by assuming:

$$prob( p_i = p_{i0}) = F(p_{i0}^U) - F(p_{i0}^L) \tag{9}$$

where $p_{i0}^U$ denotes the mid-range from $p_{i0}$ to the risk at the next upper rung and $p_{i0}^L$ denotes the mid-range from $p_{i0}$ to the risk at the next lower rung on the risk ladder. Multiplication of [8] or [9] over the appropriate respondents, i.e., those who provide a point estimate versus a range estimate, yields the sample likelihood function, which is then maximized.

4.0 Estimation results and discussion

Based on the discussion reported in Section 2, several model specifications for the median and variance of perceived risk were estimated. Table 3 reports the results from two of the most informative models. The top portion of Table 3 shows the factors affecting the median parameter of the distribution, $\mu$, whereas the lower portion shows the factors
affecting the uncertainty measure $\sigma$. All parameters are estimated using a single likelihood function for all sample respondents.

Among the variables we expected to affect the median of the distribution of perceived risk are the arsenic concentrations in the community (in ppb), an individual’s smoking status, health status, the two of these variables interacted, age, gender, treatment of drinking water at home, and current or past occupation in a risky industry. A positive sign on any of these variable coefficients indicates higher median perceived risk.

The negative sign of $Age$ in both models I and II shows that older respondents perceive less risk from arsenic exposure than a younger person, though our youngest respondents are still over 18 years in age. Further, gender ($Male$) is not statistically significant. However, we remind the reader of the fact that the risk-responders are more heavily male in composition than the group that provided no risk estimate, responses from whom could not be used in estimation of this model. Thus, concluding that gender has no significant influence on the median risk estimate must be viewed with this sample selection issue in mind. Both Models I and II also suggest that working in a Risky Occupation and being one who Treats Drinking Water do not affect the median perceived risk.

The positive and statistically significant coefficient of Arsenic Concentration, measured in ppb, in both models I and II implies that people who live in areas with relatively high arsenic concentrations perceive a greater risk. This result supports the hypothesis that people understand the connection between dose and risk. Figure 2 further examines this: it compares the estimated risk distribution for those consuming 10 ppb arsenic to those
consuming 100 ppb, with all other significant model variables held at the sample average. The distribution for high concentrations (100 ppb arsenic) exhibits dramatic risk skew, with a median perceived risk of 1390 deaths per 100,000. Median perceived risk is much lower for the 10 ppb case, at roughly 71 per 100,000.

Health Status alone is not significant at conventional levels. Smoking status enters the median model in two ways: directly as an intercept shifter (Current Smoker and Former Smoker) and indirectly through an interaction between smoking status (Current and Former) and Health Status. A somewhat surprising result is that Current Smoker x Health Status and Current Smoker variables are not significant, suggesting that all else equal, current smokers and those who never smoked have the same median risk perceptions. Even more surprising is the joint significance of Former Smoker x Health Status and Former Smoker variables. This functional form suggests that the relationship between median perceived risk and smoking varies with the subject’s current health status for former smokers only. Accordingly, those who never smoked and current smokers have a unique intercept equal to the constant term, but former smokers each have five intercepts that vary with the self-reported health measure so that as health deteriorates, the median perceived risk increases.

We do not have a priori expectation about variables that may affect the variance, thus we tried a variety of models that controlled for objective sources of uncertainty and psychosocial factors including the estimated range of arsenic concentrations in the water in the subject’s local water supply, education, gender, age, and indicator variables representing the five locales surveyed. Two models are reported at the bottom of Table 3. In model I, none of the variables are significant, but this may be evidence of multicolinearity.\(^x\) We know
that since the arsenic concentration range does not vary with locale, the *Albuquerque*
indicator and *Arsenic Concentration Range* are highly correlated. Also, a likelihood ratio
test indicates that the coefficients of the health interaction variables are not significantly
different from each other. Based on these observations, we estimate Model II, where the
interaction variables are combined giving the new variable *Ever Smoker x Health Status* and
*Arsenic Concentration Range* is dropped from the model.

Table 4 gives the estimated variances by locale, health status, and smoking status.
The *Albuquerque* indicator and *Ever Smoker x Health Status* variable are significant in the
new model. Accordingly, subjects living in the Albuquerque area have more precise
subjective risk estimates than those in other locales. The new arsenic regulation had received
significant attention in the local press just prior to our survey. It may well be that this
extensive media coverage of arsenic issues made respondents in the Albuquerque more
certain about arsenic risks, but we have no data on media coverage in the other locales to
adequately test this hypothesis (see the discussion on media coverage of arsenic problems by
Bell *et al.* 2008)

The interaction between health status and any smoking history is positive and
significant. The significance of the interaction variable suggests that information given to
subjects about the synergistic effects of smoking and arsenic induced uncertainty, especially
for those who are currently in poor health. The amount of uncertainty increases as the health
of the subject deteriorates, so that current and former smokers in poor health exhibit the most
uncertainty about arsenic risks. Current and former smokers in excellent health living in
Albuquerque are the most certain about risks. Those who have never smoked fall between
these two groups.

As noted previously, an increase in the variance of the distribution leaves the median the same, but shifts the average risk, \( \bar{p} \), to the right as the distribution becomes increasingly right skew. Table 5 gives the estimated average risk as a function of health status, smoking behavior, and locale for the average aged subject (Age=51) exposed to arsenic concentrations of 50 ppb. Recall that the information booklet stated an objective expected mortality rate of 0.01 for people exposed to this concentration for 20 or more years, with the risk double that for smokers and those in risky occupations. The results suggest that some groups tend to perceive risks lower than the objective assessments and others somewhat higher. For example, never smokers in Albuquerque perceive a risk of \( \bar{p} =0.007 \), roughly 30% lower than the objective risk whereas never smokers in other locales are closer to the mark with an average risk of \( \bar{p} =0.009 \). While unhealthy former smokers tend to overestimate the risk (for example, unhealthy former smokers in other locales have an average risk of \( \bar{p} =0.052 \)), current smokers tend to underestimate the compounding effect of smoking and arsenic consumption, as evidenced by the relatively low average risks ranging from 0.011 to 0.017. Still, current smokers have average risks higher than those who have never smoked.

It is tempting to compare the results we obtain for smokers to other smoking studies, but the reader should bear in mind that the elicited risks here pertain to arsenic, not directly to dying from smoking. However, our empirical results are consistent with some smoking studies. For example, Smith et al. (2001) found that smokers, former smokers and
nonsmokers update longevity expectations (mortality risks) differently after experiencing a health shock such as a heart attack.

5.0 Conclusions and further research

Many, if not all, mortality risks lack precision and are therefore best modeled as having some degree of uncertainty. Nevertheless, few behavioral studies of choice in the presence of risk explicitly incorporate uncertainty, presumably because it is computationally complicated. Instead, most risk-related studies in economics tend to rely upon a simple point estimate of risk for use in behavioral models, or they fall back on a single estimate of central tendency. This paper provides a computationally straightforward method for estimating the perceived risk distribution of mortality risks from arsenic in drinking water using standard risk-elicitation methods. The empirical model is parameterized to include the factors that influence both the median and variance of the perceived risk distribution, thus allowing us to estimate a model that allows for individual heterogeneity in the risk distribution.

A notable finding is that subjects conveyed average risk estimates that were not wildly out of line with objectively-assessed risks given to all respondents in the information booklet. It is possible that because drinking water is a common activity and people who live in arsenic-contaminated areas know the issues, this result is consistent with the idea that availability helps people understand arsenic-related risks (see Kahneman et al. 1982; Sjöberg and Engleberg 2010; or the recent discussion in Fischhoff 2009). Note however, we do not observe evidence of upward bias as some suggest will be found. The model revealed that perceived risk is positively associated with arsenic exposure levels and related individuating factors that should affect perceived likelihood of mortality. Even though all of our
respondents received the same information about the science of arsenic risks, risk perceptions differed across people in several ways.

In particular here, respondents’ smoking habits are among the strongest influences that affect the average individual perceived risk: all else equal, unhealthy, former smokers believe their arsenic risks to be higher than even people who currently smoke. Current smokers recognize their increased risk over those who never smoked, but still understate the average risk, as compared to the science-based estimates of risk provided in the information brochure. The key point is that while everyone slightly understates risks, the smokers in our sample appear to understand the relative mortality risks of arsenic in drinking water.

Contrary to the frequently employed empirical assumption of a zero variance in risk, we find evidence that the variance of the perceived risk distribution is non-zero. Subjects living in Albuquerque, where arsenic risks were well-publicized just prior to the survey, had more precise risk estimates than those in the other locales. And again, smoking status played an important role in the precision of the risk estimates, with unhealthy current and former smokers having less precise beliefs than those who never smoked.

The approach used in this paper has several important implications for scientists and policy makers. First, it is clear that the problem of communicating mortality risks that are scientifically uncertain needs to be better addressed and not ignored. This is not an easy task: researchers face a trade-off between trying to hold respondent’s attention for longer periods of time to obtain more clarity regarding the risks faced by specific types of individuals, at the risk of losing respondents altogether via survey non-response.
Our respondents did respond to risk information differently, supporting the notion (see Connelly and Knuth 1998) that if at all possible, risk communication should be specifically tailored to certain types of people rather than a one-size fits all approach. Indeed, the recently released CDC cancer tables break down risks by age, gender, race, and smoking habits, allowing people to assess their risks with greater accuracy and, presumably, less ambiguity than when only one average population risk estimate is given to them (Woloshin et al., 2008).

The next important step to take is to link the risk model to a behavioral model of averting behavior and/or willingness to pay for a change in risk. Other researchers [e.g. Curley and Yates 1985, Riddel and Shaw (2006), and Cameron (2005)] have typically found that when individuals make choices, they act as if they want to avoid ambiguity. For example, individuals who must make an important medical decision may seek opportunities to avoid any uncertainty pertaining to outcomes.

A recent paper by Jakus et al. (2009) finds that perceived contamination risks lead to higher expenditures on bottled water. But that analysis uses the simple stated risks and could be improved upon allowing uncertainty to play a role in the choice model. The most elegant solution would be to jointly model the risk distribution and the decision to consume bottled water. Alternatively, the risk/uncertainty model could similarly be tied to stated willingness to pay (WTP) to reduce contamination. The option price, which is the relevant measure of WTP when risk exists, has been the topic of considerable economic literature, but few have considered its meaning outside of the traditional expected-utility framework where
uncertainty is necessarily excluded (for exceptions see Smith, 1992; Jindapon and Shaw, 2008).

As in virtually any study, several other steps can also be taken to improve on our initial line of research. We have not closely examined the issue of latency in the disease as others have (e.g., Alberini et al. 2006), though our respondents were informed that scientists believe the cancers would arise only after prolonged exposure. Latency introduces ambiguity about the length of time before cancers would occur, and it is likely that this also contributes to the variance in individual’s risk distributions. Though we have included individual demographics to allow for some heterogeneity across people in their risk and uncertainty attitudes, the model could be extended to allow for unobserved heterogeneity.

Finally, it would probably be beneficial to explore the role of information in a subject’s development of the perceived risk distribution by using a split sample design where subjects are offered different information sets. This would help evaluate the risk communication process, which many have recommended (e.g. see Fischhoff 2009, and citations within his chapter). At the extreme, one way to explore different risk communication information and its effect on elicited risks is to provide one group with almost no information, and another with the most complete information available, akin to a controlled laboratory experiment. The effect of the information content can then be tested, assuming that all other factors are controlled for that influence risk perceptions. However, we discovered in focus groups that subjects balk at offering risk estimates when they have little or no information about risks, very likely because of their lack of familiarity with arsenic risks. In contrast, available and familiar risk contexts and concepts might bolster subjects’
willingness to offer risk judgments, but these in turn may be upwardly biased, as discussed in a recent issue of this journal (Sjöberg and Engelberg 2010). Thus, a lab experiment would need to additionally control for all those factors that lead to higher estimates of risk being offered.

Bostrom and Löfstedt (2003) mention, and cite supporting work that it is not only the total amount of information that must be examined, it is whether that information is actually mentally processed (see also Morgan et al. 2001). Interestingly, these authors also mention the desire to have fields that work on risk analysis better communicate with each other, i.e. psychologists and decision theorists should better communicate with economists and vice versa. We therefore provide a reminder that a common procedure in laboratory experiments is to try to integrate real consequences from answers to questions, typically by paying subjects not only for their overall participation, but also for each important risk response they provide or task that they perform (e.g. see Harrison and Rutström 2008), which may indeed help support the processing of the risk information. Though more difficult to implement, this could perhaps be done in field surveys too (Fischhoff and Beyth-Marom 1983).

Outside the laboratory split sample designs are likely to be an expensive undertaking because more subjects must be studied to have a sufficient sample size in each subgroup. As our study unfortunately demonstrated once again, risk responses, particularly coupled with information being sought on behaviors that relate to those risks, are very demanding of the respondent, making field, telephone and mail survey efforts more difficult than other types of surveys, and thus, certain to be more costly.
References


Figure 1: Distribution of risk responses (point and mid-range, deaths per 100,000)
Figure 2. Estimated probability density function of perceived risk for areas with arsenic concentrations of 10 ppb and 100 ppb.
<table>
<thead>
<tr>
<th>Area</th>
<th>Water Source</th>
<th>Mean Arsenic Concentration (ppb)</th>
<th>Range of Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albuquerque, NM (n=54)</td>
<td>Public</td>
<td>25</td>
<td>20 – 30</td>
</tr>
<tr>
<td>Fernley, NV (n=108)</td>
<td>Public</td>
<td>40</td>
<td>No range</td>
</tr>
<tr>
<td>Oklahoma City, OK (n=80)</td>
<td>Public</td>
<td>17.5</td>
<td>14 – 21</td>
</tr>
<tr>
<td>Outagamie County, WI (n=55)</td>
<td>Private, tested</td>
<td>3.84</td>
<td>No range</td>
</tr>
<tr>
<td>Outagamie County, WI (n=43)</td>
<td>Private, tested</td>
<td>—</td>
<td>5 – 105</td>
</tr>
<tr>
<td>Appleton, WI (n=5)</td>
<td>Private, tested</td>
<td>6.9</td>
<td>No range</td>
</tr>
<tr>
<td>Appleton, WI (n=8)</td>
<td>Private, tested</td>
<td>—</td>
<td>5 – 105</td>
</tr>
</tbody>
</table>
**Table 2**: Means and Standard Deviations of the Model Variables for Subjects Giving Point Estimates, Range Estimates, and Those Giving No Estimate of Risk (No Response)

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>DESCRIPTION</th>
<th>POINT (n=201)</th>
<th>RANGE (n=96)</th>
<th>ALL RESPONDERS (n=297)</th>
<th>NO RESPONSE (n=56)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>years</td>
<td>50.70</td>
<td>16.08</td>
<td>50.90</td>
<td>16.11</td>
</tr>
<tr>
<td>Male =1 if male</td>
<td>estimated concentration of arsenic</td>
<td>61.2%</td>
<td>48.9%</td>
<td>56.7%</td>
<td>49.6%</td>
</tr>
<tr>
<td>PPB</td>
<td>in local water supply, parts per</td>
<td>25</td>
<td>14</td>
<td>24</td>
<td>14</td>
</tr>
<tr>
<td>Treat Drinking Water</td>
<td>1 = subject treat drinking water at home, 0 otherwise</td>
<td>52.2%</td>
<td>50.1%</td>
<td>51.6%</td>
<td>50.0%</td>
</tr>
<tr>
<td>Risky Occupation</td>
<td>=1 if subject ever worked in a risky occupation</td>
<td>26.4%</td>
<td>44.2%</td>
<td>26.1%</td>
<td>44.0%</td>
</tr>
<tr>
<td>Former Smoker</td>
<td>=1 if the subject has quit smoking</td>
<td>35.3%</td>
<td>47.9%</td>
<td>32.9%</td>
<td>47.0%</td>
</tr>
<tr>
<td>Ever Smoker</td>
<td>=1 if the subject ever smoked</td>
<td>49.3%</td>
<td>50.1%</td>
<td>45.6%</td>
<td>49.9%</td>
</tr>
<tr>
<td>Health Status</td>
<td>graded 1 (excellent) to 5 (poor)</td>
<td>2.21</td>
<td>1.00</td>
<td>2.18</td>
<td>0.96</td>
</tr>
<tr>
<td>Education</td>
<td>graded 1 (no high school) to 7 (advanced university degree)</td>
<td>4.56</td>
<td>1.54</td>
<td>4.53</td>
<td>1.63</td>
</tr>
</tbody>
</table>
### Table 3: ML estimation of median and variance of perceived risk

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model I</th>
<th>Model II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>P-value</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.062</td>
<td>0.000</td>
</tr>
<tr>
<td>Age</td>
<td>-0.004</td>
<td>0.091</td>
</tr>
<tr>
<td>Male</td>
<td>0.018</td>
<td>0.821</td>
</tr>
<tr>
<td>Arsenic Concentration</td>
<td>0.011</td>
<td>0.000</td>
</tr>
<tr>
<td>Treat Drinking Water</td>
<td>-0.007</td>
<td>0.931</td>
</tr>
<tr>
<td>Risky Occupation</td>
<td>-0.033</td>
<td>0.726</td>
</tr>
<tr>
<td>Current Smoker</td>
<td>0.368</td>
<td>0.282</td>
</tr>
<tr>
<td>Former Smoker</td>
<td>-0.594</td>
<td>0.004</td>
</tr>
<tr>
<td>Health Status</td>
<td>-0.016</td>
<td>0.773</td>
</tr>
<tr>
<td>Health Status x Current Smoker</td>
<td>0.008</td>
<td>0.954</td>
</tr>
<tr>
<td>Health Status x Former Smoker</td>
<td>0.224</td>
<td>0.011</td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.564</td>
<td>0.000</td>
</tr>
<tr>
<td>Arsenic Concentration Range</td>
<td>0.005</td>
<td>0.746</td>
</tr>
<tr>
<td>Health Status x Current Smoker</td>
<td>0.074</td>
<td>0.163</td>
</tr>
<tr>
<td>Health Status x Former Smoker</td>
<td>0.059</td>
<td>0.152</td>
</tr>
<tr>
<td>Health Status x Ever Smoker</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Albuquerque</td>
<td>-0.275</td>
<td>0.129</td>
</tr>
</tbody>
</table>

Log-likelihood: -654.306 -654.390
Table 4. Estimated Variance, $\exp(Zb)$, by Locale, Health Status, and Smoking Status

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (excellent)</td>
<td>0.459</td>
<td>0.583</td>
<td>0.484</td>
<td>0.614</td>
</tr>
<tr>
<td>2</td>
<td>0.459</td>
<td>0.583</td>
<td>0.514</td>
<td>0.652</td>
</tr>
<tr>
<td>3</td>
<td>0.459</td>
<td>0.583</td>
<td>0.546</td>
<td>0.693</td>
</tr>
<tr>
<td>4</td>
<td>0.459</td>
<td>0.583</td>
<td>0.580</td>
<td>0.736</td>
</tr>
<tr>
<td>5 (poor)</td>
<td>0.459</td>
<td>0.583</td>
<td>0.616</td>
<td>0.781</td>
</tr>
</tbody>
</table>

Table 5. Average perceived risk as a function of health, smoking, and locale

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (excellent)</td>
<td>0.007</td>
<td>0.009</td>
<td>0.007</td>
<td>0.004</td>
<td>0.008</td>
<td>0.011</td>
</tr>
<tr>
<td>2</td>
<td>0.007</td>
<td>0.009</td>
<td>0.008</td>
<td>0.011</td>
<td>0.009</td>
<td>0.012</td>
</tr>
<tr>
<td>3</td>
<td>0.007</td>
<td>0.009</td>
<td>0.014</td>
<td>0.020</td>
<td>0.009</td>
<td>0.013</td>
</tr>
<tr>
<td>4</td>
<td>0.007</td>
<td>0.009</td>
<td>0.024</td>
<td>0.330</td>
<td>0.010</td>
<td>0.015</td>
</tr>
<tr>
<td>5 (poor)</td>
<td>0.007</td>
<td>0.009</td>
<td>0.039</td>
<td>0.052</td>
<td>0.111</td>
<td>0.017</td>
</tr>
</tbody>
</table>
Appendix: Risk Questions and Additional Materials

The survey implementation strategy again involves an information brochure sent to all respondents in the mail, and a telephone survey given to respondents after their receipt of the information brochure to collect the final information needed for the study.

The information brochure depicted the risk ladder. Following this, respondents are provided with the following text:

YOUR THOUGHTS ABOUT RISKS

In this section, we will ask you several questions about what you think the cancer risks are for different people in your household. Consider the possibility of arsenic being in your drinking water, and key factors such as age, smoking behavior, and water consumption.

We would like you to express your best estimate of risks below. There are no right or wrong answers to these questions. We know that many people who are uncertain about these risks cannot give one single value for the risk they face and prefer to provide a range from low to high. That is okay.

For each question, please make either a single mark on the ladder indicating that you have a reasonably clear idea of what the risk is OR, if you are not sure about the precise value, you can make two marks where the lowest mark indicates the lowest value you think the risk could be and the highest mark indicates the highest the risk could be.

--------------------------------------------------------------------------------------

Risk question 1. Think about your risk of dying from drinking water with arsenic in it at the current levels you have, for the past twenty years.

- If you are certain about the risks, please put a single mark on the right side of Risk Ladder 1 (next page) indicating your own of risk of dying, if you feel you can do so.

OR

- If you are uncertain about the risks, please place two marks on the right side of Risk Ladder 1 (next page), where the lowest mark indicates the lowest value you think the risk could be and the highest mark indicates the highest the risk could be.
The text used by the telephone interviewer in the final phone survey is below:

**D. YOUR THOUGHTS ABOUT RISKS**

As you can see on the information brochure we asked you to make marks with the risks from arsenic in your drinking water that you thought were relevant to you.

Next, can you go ahead and indicate the marks that you made on the ladder on the brochure for us now?

24. Adult risks:

Certain ________

Uncertain – between ________ deaths and ________ deaths

________________________________________________________________________

The entire information booklet sent to respondents is available as supplemental material, or by request of the authors, as is the entire final survey questionnaire.

Note that Jakus et al. (2009) also presents the risk ladder used in the study.
Our formal modeling of a point estimate of risk and its variance represent an improvement upon the approaches used by other researchers such as Cameron (2005) and Riddel and Shaw (2006). In Cameron’s study of climate risk, the distribution of perceived risk was elicited simply by asking for the quartile range surrounding the point estimate. Riddel and Shaw used the estimated sample variance to determine the distribution of an individual’s perceived risk.

While the sample is not representative of the United States as a whole, it was constructed to be representative of the types of people and communities facing risks associated with arsenic concentrations above the new EPA standard. It does exclude those on public systems who do not pay for their water: most likely, renters.

The phone/mail/phone procedure is demanding of respondents but was selected for several reasons. First, a mail-only survey allows no chance for clarification of potentially difficult risk concepts by the trained telephone survey team. Pre-tests of the information booklet and telephone survey allowed us to develop scripted responses to commonly asked questions. Second, a telephone-only survey would be unlikely to communicate adequately risk information more easily conveyed using visual aids via the informational booklet. Finally, laboratory experiments are attractive settings in which to explore methods of risk elicitation, such as indirect methods that rely on researchers’ inferences from risky choices that people make. However, our interest here was in a field study of a larger sample of people actually exposed to arsenic levels that violated the new standard.

Those receiving water from a public water system were provided with the mean and range of the arsenic concentration as measured by the local utility. Those on private wells that had not been tested were provided with the range of concentration known to exist in their community.

The first round recruiter survey had a response rate of 31%. Some 565 of these respondents agreed to participate in the second round survey, but second round responses
were obtained from only 353 (47\% of the original first round sample). Our attrition rate (37\%) is in line with others who have used the telephone-mail-telephone strategy.

\[ vi \] Presumably, the high degree of uncertainty in the arsenic dose-response relationship implies high subjective variance estimates, at least in comparison to relatively well-known mortality risks such as smoking. Our goal in this research was to elicit risk distributions based on current information and attitudes rather than formally testing the relationship between objective and subjective uncertainty.

\[ vii \] Riddel used a similar elicitation methodology as the current study. She allowed subjects who were confident in their assessment of risk to provide a point estimate whereas those who were less certain about risks stated an upper and lower bound.

\[ viii \] We also estimated models that included indicator variables for the education levels, but the model consistently rejected any risk-education relationship. Another specification included indicator variables for locale in both the median and variance. Only the \textit{Albuquerque} indicator variable was statistically significant.

\[ ix \] We thank a reviewer who raised this important point.

\[ x \] A specification suggested by a reviewer included an indicator variable that takes a value of 1 if the region provides a range of arsenic risks, 0 otherwise. This variable was not statistically significant.

\[ xi \] The log-likelihood test is:

\[
H_0 : \beta_{\text{arsenic range}} = 0 \text{ and } \beta_{\text{HealthStatus x Current Smoker}} = \beta_{\text{HealthStatus x Former Smoker}} \\
H_A \text{ at least one of the constraints does not hold} \\
\lambda = -2(-654.4 - (-654.3)) = 0.2 < \chi^2_1 = 3.84 \\
\text{Inference: Can't reject the null hypothesis.}
\]