Valuing lives saved from safer food:
A cautionary tale revisited

Jason F. Shogren
University of Wyoming
jramses@uwyo.edu

Tommy Stamland
Norwegian School of Economics
and Business Administration
Tommy.Stamland@nhh.no

May 2007

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Portland, OR, July 29-August 1, 2007

Prioritizing risk reduction strategies to maximize net benefits requires information on the monetary value people assign to safer food (Roberts, 1987). The challenge is that individual behavior reflects more than just unobserved preferences for risk reduction. Behavior also reflects each person's unobserved and potentially unique food risk and ability to reduce this risk privately, i.e., consumption patterns, food preparation, cleaning efforts, and so on. People with a low valuation of collective risk reduction may seem to tolerate greater risk, but they may instead have already used their unobservable skill to reduce the risk themselves (Shogren and Stamland, 2002). Food selection alone is unlikely to reveal perfectly these two characteristics because multi-dimensional heterogeneity exists in the population. In such cases, we propose using use a General Method of Moments (GMM) framework to separate risk reduction activities from risk preferences (Shogren and Stamland, 2006).

Our promotion of GMM rests on the presumption that people can make food choices based on the differentiated risks posed by different foods and gauge the appropriate responses to these risks. The open question, however, is whether people actually do differentiate low-probability food risks and respond accordingly and consistently. In experimental auctions designed to value safer food, Hayes et al. (1995) found evidence of *surrogate* bidding (also called *embedding*, *part-whole bias*, or *insensitivity to scope*). Surrogate bidding exists when behavior—either risk perceptions or values—for a specific good reflect general preferences for a phenomena rather than for the specific good in question; or when perceptions and values are insensitive to changes in the quantity or quality of the good. Hayes et al. compared the bidding behavior from foodborne pathogen treatments to the bids in a treatment that combines the risks of all the pathogens—a 1 in a 46,000 chance of illness per meal from at least one of the five pathogens. They observed surrogate bidding for reduced health risk: bids for a cluster of pathogens were indistinguishable from bids for specific pathogens.

Using a contingent valuation survey, Hammitt and Graham (1999) reproduced the Hayes et al. study, and found the same insensitivity to probability. In addition, Bateman et al. (1997) found similar results for restaurant meals. They used an incentive compatible mechanism to auction off vouchers for parts of a restaurant meal. They elicited values for the parts and the whole were elicited. They observed the sum of the parts exceeded the whole, again supporting the idea that values for food seem to reflect general preferences.

Multi-dimensional heterogeneity and surrogate bidding both present a challenge to whether we can estimate a consistent value of statistical life for safer food. First, one should separate out unobservable risk preferences from risk-reduction skill assuming differentiated responses to risk; second, one should test for whether the perceptions of risk and the responses to risk are rational to begin with. We can use the GMM method to address the first task of identification of risk preferences; and we can use internal consistency checks to address the question of differentiated responses and reactions to risk. Here we describe the initial results on rational risk valuation for safer food using data from Wave I of the University of Wyoming Food Web Diary project. We designed our UW Food Diary to provide the data needed to use a GMM framework to obtain improved valuation estimates in the face of multi-dimensional individual heterogeneity.[1]

The UW survey captures the idea that each respondent has an idiosyncratic mortality risk from each risk source, an idiosyncratic ability to reduce each of risks, and his or her own value of statistical life. We allow consumers to be heterogeneous in several observable and unobservable dimensions: tastes, budgets, base level risks, abilities to reduce risk, and their willingness to pay for risk reduction. Given this heterogeneity, rational people should choose different consumption levels based on different levels of risk. We test for rationality by asking several questions in the diary about risk perceptions and responses to risk. We create two tests of internal consistency—identical risk perception questions for three meats, and an identical risk reduction asked in two ways. Overall, even for a relatively familiar commodity like food, our results do not contradict the cautionary tale told by those concerned with surrogate behavior (e.g., Kahneman and Knetsch, 1992). On average, we observe the same question generated different responses (identical risk but different responses), and that different questions generated the same answer (insensitivity to different risks posed by pork, beef, and chicken).

Design of Survey

The general survey design consisted of five parts: general health status; risk taking behavior, knowledge of food safety and risk reduction actions; awareness of effectiveness to reduce the risks of food-borne illnesses; and risks perceptions of food-borne illnesses. We also asked for socio-demographic information. First, we asked questions to learn about the general condition of their health: Q1. Are you (or your partner) currently pregnant? Q2: Is there anybody smoking in the family? Q3. Are you exposed on a regular basis (daily/weekly) to second-hand smoke? Q4. Is there exposure to second-hand smoke inside your household; Q5. Do you consume more than two alcoholic beverages per day? Q12. How would you rate your physical health compared to others your age and gender? Q18. Does anybody in your family have specific dietary needs due to a medical condition?

Second, we asked questions about a respondent's general effort to self-protect: Q6. Imagine you will be in a vehicle 10 times; how many times would you say you would: (a) wear a seatbelt as a driver or a passenger; and (b) drive more than 5mph over the speed limit; Q7. Do you have a smoke detector in the house? Q8. Do you change your batteries in your smoke detectors at least once every year? Q9. Do you have a carbon monoxide detector in the house? Q10. Do you have a First Aid kit in the house? Q11. Do you have a fire extinguisher in the house?

Third, we asked one multi-part question about food preparation: Q33. Out of 10 meal preparations for the relevant food, how frequently do you:

a. Wash your hands with hot soapy water before handling food?
b. Wash your hands with hot soapy water after handling raw meat products?
c. Wash utensils and surfaces immediately with hot soapy water where meals are prepared?
d. Wash utensils and surfaces with hot soapy water after preparing each food item and before you go onto the next food item?
e. Wash vegetables and fruit?
f. Cook meat products to a safe temperature recommended by health experts?
g. Refrigerate leftovers within two hours of preparation?
h. Use non-expired food?
i. Use a separate cutting board for raw meat products and other non-meat items?
j. Use a meat thermometer?
k. Use plastic or other non-porous cutting boards?
l. Separate raw meat, poultry, and seafood from other foods in your shopping cart?
m. Separate raw meat, poultry, and seafood from other foods in your refrigerator?
n. Place cooked food on a plate that previously held raw meat, poultry or seafood?
o. Cook eggs until the yolk and white are firm?
p. Use recipes in which eggs remain raw or only partially cooked?
q. Cook fish until it is opaque and flakes easily with a fork?
r. Cover microwave food, stirring and rotating it?
s. Bring sauces, soup and gravy to a boil when reheating?
t. Defrost food at room temperature?

Fourth, we are interested in each person's perception of the effectiveness of the four key risk reduction methods for foodborne pathogens—washing, separating, cooking, and prompt storage. We asked four specific questions (for beef, pork, and chicken). For each question, the respondent checked off one of the intervals from (91-100%); (81-90%),…(1-10%). Q43. How effective is washing one's hands, utensils, and food before a meal in reducing the risk of food-borne illness from this food? Q44. How effective is separating raw meat/poultry/seafood from other foods and using a different cutting board for raw meat products in reducing the risk of food-borne illness from this food for a meal? Q45. How effective is cooking food to proper temperatures in reducing the risk of food-borne illness from this food for a meal? Q46. How effective is prompt storage and refrigeration in reducing the risk of food-borne illness from this food for a meal?

Next, we wanted to understand perceptions toward the change in effectiveness if the responded *cut in half* (or *doubled*, depending on survey) the frequency of washing, separating, cooking, and prompt storage, again for the three meats. For each question, the respondent circled an integer value ranging from 5 (large decrease) to -5 (large increase); e.g., Q48. If you cut in half the number of times you washed your hands, utensils, and food before a meal, how much would your risk for each type change?

Fifth, we are interested in each person's risk perception about foodborne illness. We asked two questions:

Q21. Please mark the point that you think best represents how frequently A TYPICAL AMERICAN can be expected to suffer a food-borne illness in any given year.

How Often	Frequency (In Terms of Number of People)
twice every week	(1 in 10 people)
once every month	(1 in 100 people)
once every year	(1 in 1,000 people)
once every 10 years	(1 in 10,000 people)
once every 100 years	(1 in 100,000 people)
once every 1,000 years	(1 in 1,000,000 people)
once every 10,000 years	(1 in 10,000,000 people)
once every 100,000 years	(1 in 100,000,000 people)

Q22. Please mark the point that you think best represents how frequently YOU can be expected to suffer a food-borne illness in any given year.

How Often	Frequency (In Terms of Number of People)
twice every week	(1 in 10 people)
once every month	(1 in 100 people)
once every year	(1 in 1,000 people)
once every 10 years	(1 in 10,000 people)
once every 100 years	(1 in 100,000 people)
once every 1,000 years	(1 in 1,000,000 people)
once every 10,000 years	(1 in 10,000,000 people)
once every 100,000 years	(1 in 100,000,000 people)

Implementation of Survey

Funded by the United States Department of Agriculture through the Economic Research Service, Knowledge Networks (KN) implemented the survey. The study design consisted three waves of data collection and two interventions between the three waves. Each wave of the data collection lasted for 14 days. In these 14 days, respondents were instructed to collect their household's grocery shopping receipts. After the first wave of data collection (Wave I: 1274 fielded; 923 completed; 72% response rate), respondents were sent the first intervention survey. Respondents were instructed to visit a website about ten least wanted bacteria present in foods.[2] The second wave took place subsequently (II: 905 fielded; 800 completed; 88% response rate). After the second wave, respondents were sent the second intervention survey. They were instructed to visit a different website about food safety.[3] The third wave of data collection took place (III: 774 fielded; 703 completed; 91% response rate).

The sample of the study was restricted to the overlapping panelists between KN's web-enabled panel and the National Shopper Lab (NSL) panel. This sample design was to allow analysis of the UPC data collected at the grocery stores where respondents used their NSL card to shop.[4] The following table displays the number of panelists invited for the survey and the number of completed interviews for each wave of the data collection. The survey instrument averaged approximately 19 minutes. Each respondent was awarded 5000 bonus points (an equivalent of $5) for their participation in each survey and collection of their household's grocery store receipts. The survey questionnaires are available on request.

Wave I Results

We create six indexes that summarize a subject's responses to related sets of questions; and three measures of risk perception. First, we create a health index, *HI*. Several questions pertain to the respondent's health and susceptibility to illness. We convert the answers to numeric values. For instance, if the answer to question 1 is 'yes', we define q1=1, whereas q1= -1 if the answer is ‘no’. If the question is not answered, we define q1=0. The corresponding is done with questions 2-5, 12, and 18. Based on these numerical responses, The *health index*, *HI*, is:

(1) HI = q1+q2+q3+q4+q5+q12+q18.

The questions are formulated so ‘yes’ implies poor health or higher susceptibility to illness—higher values of the health index indicate greater sensitivity to health risk.

Second, the survey respondents report in questions 13 and 14 their height and weight which we combine into the *body mass index, BMI*:

(2) $BMI = \text{weight} / \text{height}^2$,

where the weight is measured in kilograms and the height in meters.

Third, we create a measure of the respondent’s effort in self-protection through various activities that reduce risk. We have two questions that already require quantitative responses, namely the two sub-questions in question 6 pertaining to how many times, out of ten times, the respondent wears a seat belt, and adhere to the speed limit, when driving (riding) a car. The *risk index*, *RI*, is calculated as

(3) RI = q6_wear/10 – q6_drive/10 + q7 + q8 + q9 + q10 + q11 + q20 + Q24,

where Q24=1 if the response to question 24 is greater than 0, Q24=0 otherwise. The first two terms in the expressions on the right hand side of (3) indicate we use the fraction of times the respondent is wearing a seatbelt, and adhering to the speed limit, as components of the index.

Fourth, we create a food preparation index from the responses to Question 33. This question inquires about numerous activities carried out before, during, or after food preparation that may influence the heath risks posed by the prepared food. The *food preparation index*, *FPI* is calculated as the sum of the reported number of times, out of ten meal preparations, the specific activity is carried out:

(4) FPI = q33a+q33b+…+q33t

Fifth, the survey asks how effective various activities are perceived to be in reducing food borne health risks posed by three categories of food: beef, pork, and chicken. We create from the responses the following *self-protection effectiveness index*, *SPEI*:

(5) *SPEI*= q43_beef+q43_pork+ q43_chicken+…+ q46_beef+q46_pork+ q46_chicken

Questions 43-46 are formulated so that the lowest response, 1, indicates that the activity is very effective, whereas the lowest response indicates that the activity is not effective. Larger values of *SPEI* thus correspond to less effectiveness of self-protection.

Sixth, questions 48 through 51 asks, for specific activities, about the change in health risk that would result if the respondent changes the activity level. From these responses, we create an index that measures the perceived *incremental risk reduction*, *IRR*, from doubling, or cutting in half (two treatments), the number of times the risk reduction activity is carried out. The variable *amount* equals 1 if the 'cut in half' treatment was given, and *amount*=2 if the 'double' treatment was given. The responses to the questions range from 1= 'large decrease', through 6 = 'no change', to 11 = 'large increase' in risk due to the halving/doubling of the activity. We combine the responses to the questions into the index *IRR* as follows:

(6) *IRRI*= ((q48_beef-6)+(q48_pork-6)+…+ (q51_chicken-6)) x (2 x *amount* – 1)

The index is thus calculated so that it takes on negative values if doubling (halving) the activity level is associated with decreased (increased) risk, and positive values if doubling (halving) the activity level is associated with increased (decreased) risk.

Finally, questions 21 and 22, give rise to the following three risk perception variables:[1]

(7) *Personalrisk* = q21

(8) *Baserisk* = q22

(9) *Relativerisk* = *Personalrisk* / *Baserisk*

[1] We use the 'frequency value' of the response in decimal notation, for instance q21=0.001 if the response was '1 in 1000 people'.

Table 1 provides descriptive statistics about the respondents and their responses to the questions as summarized by the indexes defined above. Table 2 reports the significant correlations between the variables.

Several key points arise in the correlation matrix. First, a strong relationship exists between the health index and the body mass index. This is to be expected since both indexes are based on questions about facts and there likely is a factual relationship between health and body mass. Second, a significant correlation exists between some pairings of the indexes. The risk index (*RI*) is correlated with the food preparation index (*FPI*). Respondents with high levels of self-protection in food preparation also report more caution in other activities, such as driving. They also report that self-protection is more effective (*SPEI*). Also, respondents who believe self-protection is more effective (low *SPEI*) believe in a greater reduction of risk associated with an increase in self-protection (low *IRRI*). Given the strong similarity of the underlying perceptions asked about in the set of Q43-46 (*SPEI*) and Q48-51 (*IRRI*), we find it troublesome the correlation between the two indexes is no more than 0.115. The adjusted-R^2 is only 1.2% in a regression between *SPEI* and *IRRI*.

Third, and perhaps most notably, no evidence of correlation exists between the either of two health status indexes, *HI* and *BMI*, and any of the other indexes. The indexes, *HI* and *BMI*, are based on questions with strong factual basis, e.g., a person's weight, height, smoking, pregnancy. The risk index, *RI*, and to a lesser degree the food preparation index, *FPI*, are based in part on factual information. Some of the component questions have a weaker factual basis, such as the questions about how often, out of ten times, a person engages in a certain behavior. The 'facts' underlying such questions may be poorly remembered and may suffer from reporting errors and perhaps reporting biases. Finally, the factual underpinnings for *SPEI* and *IRRI* are probably most complex of all, and are most susceptible to reporting errors and biases.

The possibility of reporting errors and biases is hard to avoid in survey studies and we were prepared for their presence here. Our goal was to qualify the survey responses with hard data for two purposes: (1) to confront the survey responses with factual data and thereby hopefully remove some of the reporting errors and biases, and (2) to combine the survey data and the behavioral data (because neither data source contains all the relevant information) to identify each person's willingness to pay for mortality risk reduction. But both these aims require a connection between the survey data and the behavioral data. We find it disconcerting that the picture emerging from Table 2 seems to be this: there is correlation internally between the factually-oriented indexes and internally between the more subjective indexes, but there is little correlation between the two groups of indexes. There seems to be a *disconnection* between exposures to health risk and (reported) health risk reduction activities.

Fourth, this *disconnect* story is strengthened by noting the significant correlation between household income and both *HI* and *BMI*. Again, we observe correlation between factual data which appear to reflect socio-economic patterns in health and obesity. Household income is also significantly related to the general risk reduction index, *RI*, which in part is based on specific, factual questions. There is also a weaker relationship, not significant at the 1% level, with the *FPI* and *SPEI* indexes.

Fifth, the respondent's *relativerisk* is not significantly correlated with *any* of the indexes, and neither is the *personalrisk*. Interestingly, the *baserisk*, the respondent's perception of the risk exposure of *other* individuals, *is* significantly correlated (at the 1% level) with the individual's health index, *HI*, and the household income with correlation coefficients of 0.108 and -0.123. The lacking correlation between *personalrisk* or *relativerisk* with any of the other variables confirms the picture of a disconnect in the survey responses. To us, respondents seem to have vague notions of their risk exposure; so vague they correlate neither with health status, reported behavior, nor effectiveness of risk reduction.

While the evidence so far indicates a disconnection in the survey responses, the last and perhaps most consequential result we report concerns 'too high' correlation between the responses to certain questions.   Table 3 reports the correlations between the responses to the various components of Q43. The correlations are similar for Q44-46, and for Q48-51 (correlations are calculated treatment by treatment). The lowest correlation observed in any of these correlation matrixes is 0.773 while most correlations actually are above 0.9, in which the responses to Q48-51 fare worst with correlation ranging between 0.93 and 0.99 (with only one exception, a correlation of 0.88).

This strong correlation in the presence of the otherwise low to non-extant correlations is baffling. We are at a loss to explain this finding, other than the survey respondents seem to be guessing at numbers they perceive only vaguely, at best. Their guesses have a strong internal consistency between what they perceive to be similar questions, but there is no external consistency to several things we expected to be related: health status, reported behavior, reported self-protection activity, and worst of all, the responses to differently worded questions about similar underlying realities. The strong internal correlations between the responses about risks posed by beef, pork, and chicken suggest surrogate behavior, i.e., they fail the scope test as they are too flat. These results stand in sharp contrast to the lacking correlation between the responses to similar questions summarized by *SPEI* and *IRRI*.

As a final check, we examine the correlations between the responses to matched pairs of questions that form part of *SPEI* and *IRRI*. Question 43 is matched with question 48, 44 with 49, etc., and the matching is done treatment by treatment, i.e., conditional on *amount*. The contrast still exists between the internally consistent correlations in the range above 0.7 for the responses to sub-questions of the same questions; whereas the near-internal correlations with the corresponding responses to the sub-questions of the matched similar question are all below 0.177. Moreover, there is no tendency for matched sub-questions, for instance q43_beef and q48_beef, to be more correlated than non-matched sub-questions, say q43_beef and q48_chicken. The strongest observed correlation, of -0.177 (negative due to differing treatments), is between q45_chicken and q50_beef.

This strong internal consistency, and poor near-internal and external consistency, means these data are unsuitable for the GMM approach we proposed, or any other method for that matter. The internal consistency means not enough variation exists between a person's responses to the risks posed by beef, pork, and chicken which could be correlated to behavior. The lacking near-internal consistency causes us to doubt the responses represent strongly-felt perceptions. Rather, the responses appear to be 'best guesses'. Finally, the lacking external consistency indicates the responses are near orthogonal to other facts and behaviors, making them uninformative and unable to shed further light on the value of risk reduction.

In summary, the evidence from Wave I of the Food diary revealed that responses contained noise rather than information about how people perceive food safety risks and how they will respond. This implies these data cannot support reasonable VSL estimates. In the end one can question whether this result is due to the theory of rational choice, the survey design, or the respondents themselves. We conclude on a constructive note: future survey work that tries to collect data in the level of detail needed to run a GMM style model will have to find a set of different behaviors that really are distinct in the average person's mind. The challenge is to identify a framework of comparable activities that are sufficiently distinct to generate both different behavioral reactions and decided perceptions that are tied to the true nature of the activities. What we have learned ex post is that the behaviors in the risk questions have to be sufficiently differentiated, almost a day and night distinction, to make inferences about the value of statistical life given multi-dimensional heterogeneity.

References

Bateman, I., A. Munro, B. Rhodes, C. Starmer and R. Sugden (1997), "Does Part-whole Bias Exist? An Experimental Investigation." *Economic Journal* 107:322-332.

Hammitt, J., and J. Graham 1999. "Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability?" *Journal of Risk and Uncertainty* 18:33-62

Hayes, D., J. Shogren, S. Shin and J. Kliebenstein 1995. "Valuing Food Safety in Experimental Auction Markets." *American Journal of Agricultural Economics* 77:40-53.

Kahneman, D. and J. Knetsch 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management* 22:57-70

Roberts, T. 1987. "The Economic Losses Due to Selected Foodborne Diseases." Proceedings of the Ninetieth Annual Meeting of the United States Animal Health Association, pp. 336-353.

Shogren, J. and T. Stamland 2002. "Skill and the Value of Life." *Journal of Political Economy* 110: 1168-97.

Shogren, J. and T. Stamland 2006. "Consistent Estimation of the Value of Statistical Life." *Resource and Energy Economics* 28: 262-281.

TABLE 1
Summary Statistics

Variable	Mean	Std.Dev.	Min	Max
Health, *HI*	-1.4	2.7	-5.0	7.0
Body Mass, *BMI*	28.0	6.2	15.9	60.5
Risk, *RI*	4.3	2.6	-4.0	9.1
Food Preparation, *FPI*	137.9	23.6	13.0	200.0
Self-protection Effect., *SPEI*	22.9	14.3	12.0	120.0
Inc. Risk Reduction, *IRRI*	-24.1	28.4	-60.0	60.0
Household Income	64913.2	42804.2	2500.0	200000.0

TABLE 2

Correlation Matrix

	HI	*BMI*	*RI*	*FPI*	*SPEI*	*IRRI*	Income
Health, *HI*	1						
Body Mass, *BMI*	**0.178**	1					
Risk, *RI*			1				
Food Prep., *FPI*			**0.188**	1			
Self-prot., *SPEI*				**-0.25**	1		
Inc. Risk Red., *IRRI*			-0.08	-0.087	**0.115**	1	
Income	**-0.201**	**-0.091**	**0.158**	-0.073	0.064		1

Entries in bold are significant at the 1% level.
Correlations that are not significant at the 10% level are not reported.

TABLE 3

Correlation Matrix between responses to question 43

	q43_beef	q43_beef	q43_beef
q43_beef	1		
q43_pork	**0.92**	1	
q43_chicken	**0.87**	**0.89**	1

Entries in bold are significant at the 1% level.
Correlations that are not significant at the 10% level are not reported.

[1] In GMM terms, the goal is to obtain a model in which all parameters are identified. This model then yields an asymptotically unbiased estimate of the key parameter—the value of statistical life. This approach has the potential of restoring greater accuracy to the estimates of the value of statistical life by avoiding the bias inherent in single-equation approaches. This advance can come at considerable cost. Keeping the model tractable and its data requirements reasonable requires some strong assumptions. But we believe these assumptions do not invalidate the main strength of the model, which is to strip away the presumption—implicit in previous estimation approaches—that all people are identical.

[2] See http://www.fightbac.org/10least.cfm

[3] See http://www.thebody.com/fda/fsebac.html .

[4] Due to the fact that NSL participants primarily resided on the east coast and Midwest of the United States, only KN panelists in the following states were selected for the survey: Connecticut, New York, New Jersey, Pennsylvania, Ohio, Indiana, Michigan, Maryland, Washington D.C., Virginia, North Carolina, South Carolina, and Florida.