

A medical risk attitude subscale for DOSPERT

Shoshana Butler* Adam Rosman* Shira Seleski* Maggie Garcia* Sam Lee*
James Barnes* Alan Schwartz*

Abstract

Background: The Domain-Specific Risk Taking scale (DOSPERT) is a widely used instrument that measures perceived risk and benefit and attitude toward risk for activities in several domains, but does not include medical risks.

Objective: To develop a medical risk domain subscale for DOSPERT.

Methods: Sixteen candidate risk items were developed through expert discussion. We conducted cognitive telephone interviews, an online survey, and a random-digit dialing (RDD) telephone survey to reduce and refine the scale, explore its factor structure, and obtain estimates of reliability.

Participants: Eight patients recruited from UIC medical center waiting rooms participated in 45-60 minute cognitive interviews. Thirty Amazon Mechanical Turk workers completed the online survey. One hundred Chicago-area residents completed the RDD telephone survey.

Results: On the basis of cognitive interviews, we eliminated five items due to poor variance or participant misunderstanding. The online survey suggested that two additional items were negatively correlated with the scale, and we considered them candidates for removal. Factor analysis of the responses in the RDD telephone survey and non-statistical factors led us to recommend a final set of 6 items to represent the medical risk domain. The final set of items included blood donation, kidney donation, daily medication use for allergies, knee replacement surgery, general anesthesia in dentistry, and clinical trial participation. The interitem reliability (Cronbach's α) of the final set of 6 items ranged from 0.57-0.59 depending on the response task. Older respondents gave lower overall ratings of expected benefit from the activities.

Conclusion: We refined a set of items to measure risk and benefit perceptions for medical activities. Our next step will be to add these items to the complete DOSPERT scale, confirm the scale's psychometric properties, determine whether medical risks constitute a psychologically distinct domain from other risky activities, and characterize individual differences in medical risk attitudes.

Keywords: risk perception, risk attitude, DOSPERT, medical decision making.

1 Introduction

Perceptions of risk and benefit and preference for risk are important constructs in decision making under uncertainty. DOSPERT is a widely used instrument that measures conventional risk attitudes (likelihood of engaging in activities defined as objectively risky) and perceived risk attitudes (likelihood of engaging in activities perceived to be risky by the respondent) under the assumption that these constructs may differ by the domain of risk (Blais & Weber, 2006; Blais & Weber, 2009; Weber, Blais, & Betz, 2002). DOSPERT encompasses six risk domains: ethical, gambling, health/safety, investing, recreational, and social risk taking. In each domain, risk

perception, benefit perception, and risk attitude are measured using 7-point category rating scale responses to six items per domain.

Notably, although DOSPERT is one of very few measures recommended for use in assessing risks in health-related or clinical decisions in a recent review (Harrison, Young, Butow, Salkeld, & Solomon, 2005), and the only measure with extensive psychometric evidence and measures of risk perception, benefit perception, and risk attitude, DOSPERT's health/safety scale focuses on preventive safety behaviors (such as wearing a seatbelt) and does not include items that sample the types of risk activities commonly encountered in health care settings. For example, Young, et al. (2008) found no differences on the DOSPERT health/safety subscale among potential living kidney donors, potential recipients, and transplant specialists, despite marked differences in their expressed tolerance for donation-related health risks. Blais and Weber, two DOSPERT developers, also reported that the DOSPERT health/safety scale may require revision, as it appears to primarily measure variance also associated

The work reported here was funded by U.S. National Science Foundation grant SES-1023831 to Alan Schwartz. Dr. Kimihiko Yamagishi and Dr. Norimichi Hirahara contributed to the development of the initial set of items. We thank an anonymous reviewer for helpful comments on an earlier revision. Address: Alan Schwartz, Department of Medical Education (mc 591), 808 S. Wood St., 986 CME, Chicago, IL, USA 60612. Email: alansz@uic.edu.

*University of Illinois at Chicago

with the ethical domain (perhaps because several of the health/safety items reflect socially proscribed behaviors, such as unprotected sex or heavy drinking) (Blais & Weber, 2009).

The goal of this study was to develop a six-item add-on DOSPERT subscale designed to measure attitudes toward risky medical activities. This manuscript reports our work in developing the subscale and obtaining initial psychometric results for the subscale itself through three studies: cognitive interviews, an online pilot survey, and a random-digit-dialing telephone survey. These studies incorporate what have been termed the “substantive validity” and “structural validity” phases of scale development (Simms & Watson, 2007).

2 Item Selection and Cognitive Interviews

2.1 Method

Through review of literature and informal discussion among physicians and researchers, we selected 16 items to cover a wide range of medical procedures and treatment options (Table 1). The items varied in familiarity, dread, and controllability of risk (Slovic, 1987), as well as type and degree of potential benefit.

We recruited eight adults from waiting rooms at the University of Illinois at Chicago’s Family Medical Clinic and General Pediatric Clinic to participate in cognitive interviews (Willis, 2005). The goal of the interviews was to gain insight into the cognitive processes used by the respondent when answering survey questions.

Interviews were conducted over the phone and lasted 45–60 minutes. Participants were given instructions on the DOSPERT response scales and asked to rate each of the 16 items using the risk-taking, risk perception, and expected benefits scales. After participants had completed taking the survey, they were asked a series of probing questions about their answers on each item (Appendix). After participants finished the interview, they were mailed a gift card for \$20.

The focus of the analysis was to identify items that ought to be excluded from the final six-item scale. Our decisions were informed, but not dictated, by psychometric considerations; our primary concern was to identify items that were qualitatively unsuitable. Means and variances of items were computed to identify items with low variance or floor or ceiling effects. Responses to the items in the risk-taking task were analyzed for inter-item consistency with Cronbach’s alpha based on the goal of obtaining a set of items for the risk taking scale that exhibited a positive manifold. We also examined correlations between responses to each item across response tasks, to

identify items with substantially different patterns of correlation from others. Interview responses were analyzed to identify items which respondents did not comprehend or had difficulty answering.

2.2 Results

One item (Receiving a flu shot) showed a very high negative correlation of risk and benefit perception ($r = -.928$). Two items (Removing as little of the kidney as possible rather than the entire kidney when removing a kidney tumor of uncertain size; participating in a clinical trial to determine the safe dosage of a new drug) showed floor or ceiling effects and very low variance in risk taking. The first of these was also the only item that had a negative item/total correlation with other items for risk taking ($r = -.232$). Cronbach’s alpha suggested substantial inter-item consistency for risk perception (.789), risk-taking (.856), and expected benefits (.761).

Two items (Undergoing a routine colonoscopy and having a CT scan of the brain to help diagnose why you have headaches) showed the greatest misunderstandings from respondents. Many respondents confused colonoscopy with enema (e.g., explanations of the procedure by respondents included “cleaning out or removing waste from colon” and helping “digest food better and prevent infections”). Respondents were not able to explain the CT scan item in their own words.

3 Online Pilot Survey

3.1 Methods

On the basis of the cognitive interviews, we removed the five most problematic items, leaving 11 candidate items. We administered a second survey containing these items online to 30 participants using Amazon Mechanical Turk (AMT), an online labor market increasingly used for behavioral research. AMT has been shown to give results that are valid and comparable to laboratory based testing in less time and with less money (Buhrmester, Kwang, & Gosling, 2011; Paolacci, Chandler, & Ipeirotis, 2010).

We computed descriptive statistics as well as Cronbach’s alpha for each response task. In addition, we created scales for each response task by averaging item responses and explored relationships between participant gender and age on each scale. Our goal again was to characterize responses and to determine if there were items that were clearly unsuitable for use, without compromising the content validity of the scale.

Table 1: Item properties for risk-taking task (interview and online survey studies).

Item	Cognitive interviews (16 items)				Online survey (11 items)			
	Mean (SD)	Skew (Kurtosis)	Min–Max	Item-tot. corr.	Mean (SD)	Skew (Kurtosis)	Min–Max	Item-tot. corr.
1 Donating bone marrow	4.00 (2.62)	.32 (–2.2)	1–7	.347	3.63 (2.06)	.23 (–1.4)	1–7	.274
2 Donating a kidney	3.63 (2.33)	.70 (–1.1)	1–7	.188	2.23 (1.65)	1.3 (.95)	1–7	.257
3 General anesthesia	4.25 (2.49)	–.53 (–1.9)	1–7	.909	4.33 (2.35)	.43 (.83)	1–7	.371
4 Giving blood	5.63 (2.33)	–1.6 (1.2)	1–7	.340	5.57 (1.70)	.43 (.83)	1–7	.304
5 Flu Shot	4.38 (2.88)	–.16 (–2.5)	1–7	.535				
6 Daily med. for allergies	5.38 (2.00)	–.83 (–.80)	2–7	.566	4.73 (2.08)	–.70 (–.61)	1–7	.394
7 Daily med. for asthma	4.38 (2.26)	–.23 (–1.4)	1–7	.352	3.67 (2.07)	.26 (–1.2)	1–7	.485
8 Removing kidney	6.75 (0.46)	–1.4 (0.0)	6–7	–.232				
9 Back surgery	4.13 (2.53)	–.19 (–1.9)	1–7	.450	3.23 (1.79)	.35 (–1.1)	1–7	.662
10 Knee surgery	5.00 (2.33)	–.99 (–.41)	1–7	.857	3.30 (1.73)	.23 (–.85)	1–7	.511
11 Clinical trial for drug dosage	1.38 (0.74)	2.0 (3.2)	1–3	.436				
12 Clinical trial for drug effectiveness	1.63 (1.41)	2.5 (6.5)	1–5	.491	3.07 (1.89)	.23 (–1.3)	1–7	.516
13 Radiation therapy	4.75 (2.32)	–.43 (–1.4)	1–7	.357	4.77 (1.76)	–.72 (–.34)	1–7	.327
14 Colonoscopy	5.38 (2.07)	–1.6 (2.5)	1–7	.669				
15 CT scan	5.75 (2.38)	–1.7 (1.4)	1–7	.666				
16 Pain narcotics	5.50 (1.93)	–1.3 (.18)	2–7	.633	5.40 (1.65)	–1.2 (.89)	1–7	.062

3.2 Results

The AMT sample included 11 men and 19 women. Respondents were age 18–58, with median age 35.

None of the items exhibited substantially floor or ceiling effects. Cronbach’s alphas for the complete set of 11 items were 0.73 (risk-taking), 0.48 (risk perception), and 0.85 (benefit perception). Two items, donating bone marrow and taking narcotics for postoperative pain, behaved somewhat differently than other items; removing them resulted in Cronbach’s alphas of 0.75, 0.60, and 0.84, respectively. However, we felt that these differences were not strong enough for us to remove items on the basis of this pilot survey, and we retained all 11 items for the final survey phase. No differences were found between scale scores by gender or age category.¹

¹At the suggestion of a reviewer, we also considered applying a purely psychometric approach of removing items with item-total correlation less than 0.3. This would result in removing item 2 (on the basis of the cognitive interview) and, based on recomputed item-total correlations in the online survey, would further necessitate removing items 1, 4, and 16. Measures of inter-item consistency were not substantially different as a result. Because this would result in the elimination of all three items that sample medical activities in which expected benefits are purely social or psychological (i.e., donation), we maintained our original approach in order to ensure representation of this subdomain of medical activities in the telephone survey.

4 Random-digit-dialing Telephone Survey

4.1 Methods

In the final study, we administered the 11-item survey to 100 participants obtained via telephone. Phone numbers were randomly generated digits preceded by a randomly generated area code from the greater Chicago area (including outlying suburbs; area codes included 224, 312, 331, 630, 708, 773, and 847), with prefixes known to be assigned to cellular phone providers excluded. Participants were limited to English-speaking adults with no other exclusion criteria. Three attempts were made to contact telephone numbers with no answer or with answering machines (messages were not left); refusals were never recontacted. The average phone interview lasted about 10 minutes. The sample size was chosen to provide a sufficient sample for exploratory factor analysis of the items and to enable comparison of response scales among demographic groups (Fabrigar, Wegener, MacCallum, & Strahan, 1999).

Participants were first informed of the general nature of the survey and then asked if they were interested in participating. If they consented, the 11-item survey was

then administered for each scale beginning with the risk-taking scale, followed by the risk perception scale, and ending with the expected benefits scale. Participants were instructed to answer each question as if they were to find themselves in the given situation using the standard DOSPERT instructions and response scales. Participants were also asked their age, sex, and race/ethnicity for demographic purposes. Participants were mailed a \$10 gift card for their time.

The goal of the telephone survey was to examine the psychometric properties of each question in a representative urban area. Descriptive statistics were calculated, along with exploratory factor analysis to better understand the dimensional structure of the items. Factor analyses were conducted with maximum likelihood extraction and direct oblimin rotation to permit correlated factors, and the number of factors to be extracted was selected by examination of Scree plots (Fabrigar, et al., 1999). When factor analysis suggested unidimensionality or a small number of correlated dimensions, we computed Cronbach's alpha for the complete set of items and item-total correlations for each item. Finally, items were averaged into scales and analyses of covariance (ANCOVA) were conducted with sex (male/female) and race (Caucasian/African-American/Other) as independent variables and age (in years) as a covariate.

4.2 Results

We calculated contact rates and cooperation rates using the standards of the American Association for Public Opinion Research (CON3, COOP2) (American Association for Public Opinion Research (AAPOR), 2006). Contact rate 3, based on 15,817 numbers dialed to obtain 666 eligible households, was 4.2%, and reflects the random generation of phone numbers, including large numbers of non-working and non-household numbers. Cooperation rate 2, based on 100 respondents completing interviews from a total of 666 eligible households, was 15%.

The demographics of the respondents were similar to those of Cook County, Illinois. Female respondents comprised 51% of the sample. Respondents' ages ranged from 19 years old to 90 years old, and were normally distributed with a mean of 47 years and a standard deviation of 17 years. Respondents reported their race/ethnicity as Caucasian (59%), African-American (28%), Hispanic (6%), Asian/Pacific Islander (2%), or reported another race or declined to respond (5%).

Table 2 presents descriptive statistics for the items on each of the DOSPERT tasks. Item skew varied from -1.2 to 0.3 in the risk-taking task, from -0.3 to 1.4 in the risk perception task (with the exception of giving blood, which had skew of 2.5), and from -0.6 to 0.6 in the benefit perception task. Kurtosis values ranged from -2 to

2 across all tasks, with the exception of risk perception for giving blood, which had kurtosis of 7.5 . No floor or ceiling effects were evident in any of the items in any response task; respondents used the full scale.

Exploratory factor analysis of the risk-taking responses suggested a two-factor solution in which the three donation items (bone marrow, kidney, blood) loaded onto one factor and most other items loaded onto the second factor, with the anesthesia for wisdom tooth removal item not loading strongly onto either factor. Factors were moderately correlated ($r=0.36$) and inspection of the loading plots suggested that a single-factor solution would also characterize the data well.

Exploratory factor analysis of the risk perception responses suggested a two-factor solution in which the more unfamiliar procedures (surgery, clinical trials, radiation therapy, kidney donation) loaded onto one factor and most other items loaded onto the second factor, with the bone marrow donation item not loading strongly onto either factor. Factors were moderately correlated ($r=0.36$) and inspection of the loading plots again suggested that a single-factor solution would also characterize the data well.

Exploratory factor analysis of the benefit perception responses suggested a three-factor solution. The two items involving daily medication and anesthesia for wisdom tooth removal loaded onto the first factor; the three items involving donation (in which the benefit is primarily social, rather than personal) loaded onto the second factor; the remaining items loaded onto the third factor. Factors 1 and 3 were intercorrelated ($r=0.48$), but factor 2 less so ($r=-.02$ with factor 1, $r=.25$ with factor 3). Inspection of the loading plots suggested that a two-factor solution (factor 2 vs. factors 1 and 3 combined) would also characterize the data well.

Cronbach's alphas for the full set of 11 items were 0.72 (risk-taking), 0.76 (risk perception), and 0.75 (benefit perception). Item-total correlations were positive and moderate for all items. Mean risk-taking scores were negatively correlated with mean risk perception scores ($r=-0.47$, $p<.001$) and positively correlated with mean expected benefits ($r=0.54$, $p<.001$). Risk perceptions and expected benefits were negatively correlated, but not strongly so ($r=-0.29$, $p=.003$).

ANCOVA found no effects of age, gender, or race on mean risk-taking or risk perception scale scores. For benefit perception scores, older age was associated with significantly lower benefit perceptions for medical risks ($F(1,89)=241$, $p<.001$). Examination of the raw data suggests that mean benefit perceptions are relatively similar (between scale values 4 and 5) for respondents up to age 55 and then decline linearly from scale value 4 to 2.5 as respondent age increases.

Table 2: Item descriptive statistics (RDD telephone survey).

Item	Risk-taking		Risk perception		Benefit perception	
	Item-total		Item-total		Item-total	
	Mean (SD)	corr.	Mean (SD)	corr.	Mean (SD)	corr.
1 (Donating bone marrow)	4.08 (2.04)	.38	3.50 (1.94)	.24	3.27 (2.24)	.34
2 (Donating a kidney)	3.47 (2.00)	.43	4.34 (1.99)	.27	3.17 (2.09)	.39
3 (General anesthesia)	4.34 (2.32)	.16	2.98 (1.65)	.36	3.80 (2.07)	.33
4 (Giving blood)	5.51 (2.08)	.35	1.62 (1.18)	.31	4.19 (2.08)	.42
6 (Daily med. for allergies)	5.04 (2.10)	.35	2.18 (1.37)	.51	4.36 (1.64)	.41
7 (Daily med. for asthma)	4.28 (2.24)	.26	2.29 (1.42)	.55	3.95 (1.88)	.29
9 (Back surgery)	4.17 (2.10)	.48	4.45 (1.73)	.45	4.51 (1.69)	.62
10 (Knee surgery)	4.56 (2.04)	.48	3.88 (1.71)	.48	4.56 (1.81)	.48
12 (Clinical trial for drug effectiveness)	3.49 (2.19)	.44	4.91 (1.76)	.54	3.04 (1.98)	.36
13 (Radiation therapy)	5.16 (1.75)	.48	4.37 (1.84)	.46	4.66 (1.71)	.36
16 (Pain narcotics)	5.30 (1.92)	.27	3.28 (1.64)	.36	4.59 (1.72)	.45

Seven respondents indicated that giving blood was riskier than donating bone marrow or riskier than donating a kidney. Three (different) respondents gave the same response to every item in at least one task, with one responding that all items were “not at all risky”, one responding that all items were “beneficial”, and one responding that all items were “not at all risky” and “not at all beneficial”. Removing these 10 respondents from the data set and repeating the analyses did not result in substantial differences.

5 General Discussion

In these studies, we identified a set of intercorrelated items for measuring constructs related to risky medical activities within the DOSPERT framework. After eliminating five items for cognitive and psychometric reasons, the remaining items demonstrated acceptable consistency in each task (under assumptions of unidimensionality) and factor structures that suggest that although there may be more than one underlying dimension in responses to the items under some tasks (particularly in the case of benefit perception, where medical activities may benefit the actor, others, or both), these dimensions are themselves correlated. As expected, mean risk-taking scores were negatively correlated with risk perceptions and positively correlated with expected benefits.

Our goal was to develop a six-item subscale for use alongside the existing DOSPERT subscales, which imposes some unusual constraints on scale development: we explicitly are seeking a scale with exactly six items, and the same items must be evaluated under three differ-

ent tasks. Our results did not suggest strong or consistent psychometric arguments for eliminating any specific additional items from among the 11 items tested in the telephone survey, although the online survey suggested that donating bone marrow and taking narcotics for post-operative pain might be candidates for removal. Were the items not intended as a new DOSPERT domain, arguments could be advanced for retaining the complete 11-item scale, but the desire to produce a six-item scale for compatibility with other DOSPERT domains requires a further selection process. Selection is further complicated by the need for the same set of items to be used in each of the three DOSPERT tasks. This process could be conducted based on purely psychometric considerations (e.g., strictly retaining those items with the greatest evidence for unidimensionality on multiple tasks), but we chose instead to focus retaining the content validity of the scale and maximizing opportunities for other investigators to measure potentially informative variation in responses to items. As Clark and Watson (1995) note, “maximizing internal consistency almost invariably produces a scale that is quite narrow in content; if the scale is narrower than the target construct, its validity is compromised.” For example, we elected to retain the “giving blood” item, despite highly skewed distribution of perceived risk, because it represents a common medical activity for which people vary considerably in their willingness to participate and their beliefs about risk and benefit.

Accordingly, we propose a six-item subscale from the remaining nine items on the basis of ensuring that the items broadly cover the features of the activities, including their invasiveness, medical vs. surgical nature, and

Table 3: Proposed DOSPERT medical risk domain subscale items.

Item	Item-total correlations		
	Risk-taking	Risk perception	Benefit perception
M1 Donating one kidney to a patient you do not know	.32	.14	.34
M2 Giving blood	.30	.29	.38
M3 Participating in a clinical trial to determine whether a new drug is effective	.37	.42	.28
M4 Taking daily medication to relieve allergy symptoms	.25	.40	.25
M5 Knee replacement surgery to treat arthritis	.41	.41	.36
M6 Receiving general rather than local anesthesia when having a wisdom tooth removed	.23	.36	.34

differences in recipients of benefits. For example, we chose to retain only one of the two tissue donation (kidney, bone marrow) items, only one of the two surgery (knee, back) items, only one of the two daily medication (allergy, asthma) items, and only one of the two analgesic (general anesthesia, narcotics) items. We also removed the radiation therapy item for reasons exogenous to the responses; we intended at the outset to test the final subscale in Japan as well as the United States, and historical events during the study made items related to radiation likely to be a sensitive issue for Japanese citizens. Table 3 presents the proposed subscale and post hoc calculation of item-total correlations for each item based on responses in the telephone survey. Post hoc calculations of interitem consistency for this scale yield Cronbach's alphas of 0.57 for risk-taking, 0.59 for risk-perception, and 0.59 for benefit perception among our telephone survey respondents, and similar patterns of intercorrelations between the tasks as found in the 11-item scales. ANCOVA with these six items similarly found decreasing benefit perception with age. Caution should be taken in relying on these statistics, as they are based on administration of the 11-item scale, and it is possible that these six items might behave differently when presented on their own or in the context of the complete DOSPERT; this remains an area for future investigation.

This study has several limitations. As we focused on the initial development and characterization of the new subscale, considerable work remains to be done in administering the complete expanded DOSPERT to a larger population. Although we ask about reactions to risky medical activities, we follow the lead of the DOSPERT developers in that we rely on self-report and do not observe actual medical choices or measure the health status of respondents. Future studies should confirm that risk and benefit responses in the medical domain subscale are distinct from those in other DOSPERT domains, and in-

vestigate the degree to which DOSPERT domain scores predict medical decisions by patients.

References

- American Association for Public Opinion Research (AAPOR). (2006). Standard definitions: Final dispositions of case codes and outcome rates for surveys Retrieved 9/20/2008, from http://www.aapor.org/uploads/Standard_Definitions_04_08_Final.pdf.
- Blais, A., & Weber, E. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, 1(1), 33-47.
- Blais, A., & Weber, E. U. (2009). *The Domain-Specific Risk Taking (DOSPERT) scale for adult populations: Item selection and preliminary psychometric properties*. Toronto, Ontario: Defence R & D.
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk. *Perspectives on Psychological Science*, 6(1), 3-5.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309-319.
- Fabrigar, L., Wegener, D., MacCallum, R., & Strahan, E. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272-299.
- Harrison, J. D., Young, J. M., Butow, P., Salkeld, G., & Solomon, M. J. (2005). Is it worth the risk? A systematic review of instruments that measure risk propensity for use in the health setting. *Social Science & Medicine*, 60(6), 1385-1396.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5(5), 411-419.
- Simms, L. J., & Watson, D. (2007). The construct validation approach to personality scale construction. In R.

- W. Robins, R. C. Frayley & R. F. Kruger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 240-258). New York: Guilford.
- Slovic, P. (1987). Perception of risk. *Science*, 236(4799), 280-285.
- Weber, E. U., Blais, A. R., & Betz, N. E. (2002). A Domain-specific Risk-attitude Scale: Measuring Risk Perceptions and Risk Behaviors. *Journal of Behavioral Decision Making*, 15(4), 263-290.
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications, Inc.
- Young, A., Karpinski, M., Treleaven, D., Waterman, A., Parikh, C. R., Thiessen-Philbrook, H., et al. (2008). Differences in tolerance for health risk to the living donor among potential donors, recipients, and transplant professionals. *Kidney International*, 73(10), 1159-1166.

Appendix

Cognitive interview probing questions:

1. Can you explain this activity in your own words?
2. You chose (insert response) to have this procedure, what makes this (insert response) as opposed to others?
OR: How did you come to the answer of (insert response) over the other choices provided?
3. You said you consider this (insert response) risky. What makes it (insert response) risky?
4. How did you weight the benefits versus the risks when answering this question?
5. How difficult were these questions to answer? Why?