

Predicting (un)healthy behavior: A comparison of risk-taking propensity measures

Helena Szrek* Li-Wei Chao† Shandir Ramlagan‡ Karl Peltzer§

Abstract

We compare four different risk-taking propensity measures on their ability to describe and to predict actual risky behavior in the domain of health. The risk-taking propensity measures we compare are: (1) a general measure of risk-taking propensity derived from a one-item survey question (Dohmen et al., 2011), (2) a risk aversion index calculated from a set of incentivized monetary gambles (Holt & Laury, 2002), (3) a measure of risk taking derived from an incentive compatible behavioral task—the Balloon Analog Risk Task (Lejuez et al., 2002), and (4) a composite score of risk-taking likelihood in the health domain from the Domain-Specific Risk Taking (DOSPERT) scale (Weber et al., 2002). Study participants are 351 clients of health centers around Witbank, South Africa. Our findings suggest that the one-item general measure is the best predictor of risky health behavior in our population, predicting two out of four behaviors at the 5% level and the remaining two behaviors at the 10% level. The DOSPERT score in the health domain performs well, predicting one out of four behaviors at the 1% significance level and two out of four behaviors at the 10% level, but only if the DOSPERT instrument contains a hypothetical risk-taking item that is similar to the actual risky behavior being predicted. Incentivized monetary gambles and the behavioral task were unrelated to actual health behaviors; they were unable to predict any of the risky health behaviors at the 10% level. We provide evidence that this is not because the participants had trouble understanding the monetary trade-off questions or performed poorly in the behavioral task. We conclude by urging researchers to further test the usefulness of the one-item general measure, both in explaining health related risk-taking behavior and in other contexts.

Keywords: BART, CRRA, DOSPERT, monetary gambles, problem drinking, risk behavior, risky sexual behavior, risk-taking propensity, seat belt use, smoking.

1 Introduction

Risk taking is part of life, but people differ in their risk-taking propensity. Some people enjoy risky pursuits while others detest such activities. Psychologists and economists have developed various methods to measure

an individual's risk-taking propensity. This paper examines four different types of risk-taking propensity measures and tests how well they predict risk taking in the health domain, for smoking, problem drinking, seat belt non-use, and risky sexual behavior.¹

The first risk-taking propensity measure, called the Dohmen measure in this paper, was developed by Dohmen et al. (2011). It is elicited using one survey question that directly asks about risk-taking propensity: “How do you see yourself? Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Please tick a box on the scale, where the value 0 means ‘not at all willing to take risks’ and the value 10 means ‘very willing to take risks’.” While straightforward and simple to administer, the Dohmen measure has been subject to two concerns. The first concern relates to domain specificity. It is generally accepted in psychology that behavior is domain specific and that it is possible for

The authors gratefully acknowledge financial support for the study provided by the National Institutes of Health National Institute on Aging (P30AG12836, B. J. Soldo, P.I.), the Portuguese Foundation for Science and Technology, and the European Social Fund. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders. The authors thank the fieldwork coordinators Jesswill Magerman and Patricia Makuba, field supervisors Maria Vilakazi and Mantwa Mofokeng, and dedicated field workers Bridget Chiloane, Sellinah Gume, Marcia Lingwathi, Deborah Magolo, Thandiwe Mthombothi, Cleopatra Ncongwane, Nontokozo Shabangu, and Nonhlahlhla Sibanyoni. The authors also thank the Mpumalanga Department of Health for providing logistical support and clinic space for fieldworker training and study implementation.

*Centre for Economics and Finance, University of Porto. Address: CEF.UP, Faculty of Economics, Rua Dr. Roberto Frias, 4200–001, Porto, Portugal. Email: hszrek@wharton.upenn.edu.

†Population Studies Center, University of Pennsylvania and Social Aspects of HIV/AIDS and Health, Human Sciences Research Council.

‡Social Aspects of HIV/AIDS and Health, Human Sciences Research Council.

§Social Aspects of HIV/AIDS and Health, Human Sciences Research Council and Department of Psychology, University of Limpopo.

¹In this paper, we do not try to review the literature on risk measurement. We refer the interested reader to Weber et al. (2002) and Weber (2010) for an overview of different risk frameworks, to Weber and Johnson (2008) for an overview of psychological, economic, and neurological approaches to explaining risk preference, and to the decision making individual differences inventory (<http://www.sjdm.org/dmidi>) (Apelt et al., 2011) for a taxonomy of different risk attitude measures.

individuals to show different levels of risk-taking propensity depending on the context (Bromiley & Curley, 1992; Weber et al., 2002). Thus, the Dohmen measure, being a general measure, may not be able to predict risk taking in the health domain, for example. This concern is greatly mitigated by Dohmen et al. (2011), who showed the general Dohmen measure to be a robust predictor of behavior in different domains, including smoking behavior in the health domain.

The second concern with the Dohmen measure is its lack of incentive compatibility. This issue derives from a general concern that economists have about incentive-compatible measures, in which respondents are presumed to “work harder, more persistently, and more effectively” when money is aligned with performance (Camerer & Hogarth, 1999). In the case of measuring people’s risk-taking propensity, incentive compatibility is usually achieved by giving respondents choices between gambles and paying them for what they choose. Because the Dohmen question does not provide incentives to motivate respondents to answer truthfully, respondents may misrepresent their answers. Charness and Viceisza (2011), for instance, administered the Dohmen question in rural Senegal and found the resulting distribution of answers to differ significantly from that found by Dohmen et al. (2011) in Germany. The authors implied that not having aligned incentives was a reason for the discrepancy between the measures found in Senegal and in Germany, but the authors also alluded to issues related to translation difficulties, including the lack of an equivalent word for “risk” in Wolof, the main national language of Senegal.

The second risk-taking propensity measure, called the HL measure in this paper, was developed by Holt and Laury (2002). HL is the most common measure of an individual’s risk-taking propensity in the economics literature. The participant in HL makes ten separate choices between gambles, where each choice entails choosing between a “risky” gamble (where high and low payoffs are R48 and R2, respectively, in our study) and a “safer” gamble (where the high and low payoffs are respectively R25 and R20).² The probability of obtaining a high vs. low payoff in the “risky” gamble is the same as the probability of obtaining a high vs. low payoff in the “safer” gamble. This probability ranges from 0 to 0.9, in increments of 0.1, for the ten sets of choices. When the probability of the high payoff is high, such as at 0.9, the respondent has a high chance of ending up with R48 under the “risky” option and R25 under the “safer” option; in this case, only the very risk averse would choose the “safer” option. The point at which the respondent switches from the “safer” option to the “risky” option can be used to cal-

culate a risk aversion index. Because one of the choice pairs is randomly selected to be played for real after the respondent has made all 10 choices (and the respondent knows about this before making any choices), the HL measure is incentive-compatible.

Because the HL measure is derived using monetary gambles, it is unclear whether a measure derived from a monetary domain is related to risky health behavior. This question has not been widely studied. Anderson and Mellor (2008), using a sample of almost 1,000 mostly college-educated respondents in the U.S., found the HL measure to be associated with smoking, heavy drinking, and seat belt non-use—but only at the 10% significance level. Lammers (2008), using a sample of about 100 college students in South Africa, found the HL measure to be unrelated to condom use.

The third risk-taking propensity measure used in this paper is the BART (Balloon Analog Risk Task; Lejuez et al., 2002), which is a behavioral risk task performed on a computer. The task has 30 trials. Each trial begins with an un-inflated balloon on the screen. Clicking on the computer mouse makes the balloon grow bigger and earns the participant 5 cents per click. Every balloon has a different pre-set level of inflation before it bursts, and the participant does not know the number of clicks that can be made before any balloon bursts. At any point during each trial, the participant can stop clicking and bank the money earned for that trial (and the computer makes a slot-machine coin-dispensing sound). If, however, the balloon is over-inflated beyond its pre-set level, the balloon bursts and the money for that trial is gone. At the end of the task, the participant is paid all the money that is banked. The BART task, in essence, measures the actual risk-taking propensity of the individual. Because earnings feedback is given after each trial with either a casino sound when the money is banked, or an explosion noise if the balloon bursts, BART decisions are made in a “hot” psychological state (Figner et al., 2009). HL decisions, by contrast, are made in a “cold” psychological state, because the actual earnings from the HL task are revealed after all choices have been made. Dislich et al. (2010) use the term “impulsive risk taking” to describe BART and the term “reflective risk taking” to describe a task similar to the HL.

The BART has been shown to be associated with psychological measures of risk taking (such as sensation seeking and personality) and self-reported behavioral risks (including smoking, heavy drinking, drug use, sexual risk taking, gambling, stealing, etc.) (e.g., Lejuez et al., 2002). However, that BART is a good predictor of health related risk taking is not a consistent finding in the literature. Reynolds et al. (2006) found that the BART was uncorrelated with the study’s self-reported personality measures related to risk taking. Gordon (2007) found

²Our study was conducted in late 2009 when the exchange rate was 7.5 South African rand to 1 United States dollar. With R48, respondents could buy eight loaves of bread.

BART to be uncorrelated with attitudes towards risky driving. Dean et al. (2011) found that smokers did not display greater risk taking on the BART than nonsmokers. Klassen (2010) found that the BART was unrelated to any of the alcohol consumption measures examined.

Our fourth measure uses a modified version of the Domain Specific Risk Taking scale (DOSPERT), developed by Weber et al. (2002). The DOSPERT asks about one's likelihood to pursue various hypothetical but risky activities in each of five domains (financial, ethical, health/safety, social, and recreational), and then aggregates the *activity*-specific risk-taking propensity scores into *domain*-specific risk-taking propensity scores. This is in contrast to the Dohmen measure, which elicits a single global risk-taking propensity. The DOSPERT domain scores have been shown to be associated with real-life risk taking in activities within the same domain. For instance, Hanoch et al. (2006) found that smokers had significantly higher risk-taking propensity scores in the health domain than respondents in a comparison group consisting of gym members, athletes, gamblers, and investors. Zuniga and Bouzas (2005) found that the DOSPERT health/safety and recreational risk-taking scores significantly predicted blood alcohol concentrations in Mexican high-school students.

In the current study, we examine Dohmen, HL, BART, and DOSPERT measures in a population of clients at health clinics in South Africa, and test how these measures relate to health risk taking, in terms of smoking, drinking, seat belt non-use, and risky sexual practices. Our study makes a few contributions. We extend the work of Dohmen et al. (2011), by testing how the one-item measure fares in predicting unhealthy behavior other than smoking. We test the ability of the incentive-based HL and BART tasks in explaining activities in the health domain, using an implementation of the HL task that may be of interest to other researchers as it addresses concerns (e.g., Deck et al., 2010) about respondent understanding of this risk task. We also test the usefulness of the DOSPERT in its ability to predict specific risky activities within the health domain, in contrast to the usual application of the DOSPERT in explaining differences in risk taking across domains.

2 Method

2.1 Participants

Participants (N=351) were a subsample of respondents referred from another study that took place in health clinics around Witbank in the mostly rural province of Mpumalanga, South Africa. That other study recruited a systematic sample of consecutive post-HIV testing clients who visited the study clinics during a five-month period.

Every HIV positive respondent and every third HIV negative respondent was recruited, resulting in a sample with almost a 1:1 ratio of HIV positive to HIV negative clients. Participation in our study took place two months after the participants' HIV tests, and, due to higher attrition of HIV+ clients, 60% of our sample was HIV negative. Participants were reimbursed travel expenses and were paid the money earned in the various activities of the study. Data were collected between September 2009 and April 2010.

2.2 Measures

Self-reported risky behavior. We measured self-reported risky behavior for the following health risks: smoking, problem drinking (alcohol), seat belt non-use, and risky sexual behavior. Smoking was defined as currently using any tobacco products, such as cigarettes, snuff, chewing tobacco, or cigars. Problem drinking was assessed using the three question Audit-C (Bush et al., 1998), which defines problem drinking as a function of whether the respondent drinks alcohol, the number of drinks containing alcohol the respondent drinks on a typical day, and how often the respondent consumes over five (four for women) alcoholic drinks in a single sitting. Seat belt use was assessed by asking the respondent if a seat belt was used the last time the respondent sat in the front seat of a car. We flagged sexual behavior as risky if the respondent did not have a regular partner and did not use a condom the last time the respondent had sex or if the respondent was married/cohabitating and had more than one partner in the last 12 months.

Risk-taking propensity measures. We measured risk-taking propensity using the methods as described earlier in this paper (Dohmen, HL, BART, DOSPERT), with some modifications as follows:

The Dohmen measure was modified from the original 0 through 10 scale to a 1 through 7 scale.

HL was explained in great detail by the use of diagrams as well as by the use of actual buckets containing balls marked with high or low values that were shown to respondents. The random drawing of a ball to determine the actual pair (out of 10 pairs) of "risky" vs. "safer" gambles to be played for real, as well as the random drawing of one of the 10 balls from the chosen bucket, was demonstrated numerous times using different hypothetical choices. Participants understood before they made their choice that the drawing of the ball for the chosen bucket would be made afterwards and that they would receive the amount that they randomly "pulled" from the bucket. See Appendix Figure 1a for the diagram used by the interviewer to elicit the trade-offs.

The BART was administered using netbook computers, with a total of 30 trials with 5 cents (in rand) allotted

per unexploded pump. See Appendix Figures 2a and 2b for BART instructions and the diagram used to describe the task.

The DOSPERT items took the form of “If the opportunity arises, how likely do you think you will actually: *Drink heavily on weekends*”—with different risky activities inserted in the italics. After modifying items for cultural appropriateness, we implemented a limited DOSPERT, with two questions on recreational risk taking (cool off in a fast-flowing river with shoulder-deep water on a hot summer day; go rock or mountain climbing), two questions on gambling risk taking (bet a day’s income on lottery tickets or scratch cards; bet a day’s income on ifafi, umChina, cards and dice, or horse racing), and five questions on health risk taking (buy and use an illegal drug; smoke half a pack of cigarettes a day; drink heavily on weekends; sit in the front seat of a car without a seat belt; have sex with a new partner without a condom). Using these questions, we created various DOSPERT indices, including the DOSPERT-*general* (using average scores from all nine items), the DOSPERT-*health* (with only the five health-related items), and the DOSPERT-*nonhealth* (with only the recreational and gambling items).

Demographics. We collected information on age, gender, the highest level of education completed, marital status, average household monthly income, the importance of religion, and numerical reasoning using a 5-item instrument. The numeracy instrument is provided in Appendix Figure 3.

2.3 Procedure

All participants went through the same sequence of activities as follows. The supervisor first administered an economic activity that was unrelated to the present study, followed by the HL activity and then the BART activity. An interviewer then administered the DOSPERT survey, the Dohmen question, the numerical reasoning questions, and another 45 minutes of questionnaire survey (with items on HIV-related stigma, depression, etc.). Questions on actual health-related risky behavior came at the end of the survey. After the survey, the participant drew the balls to determine their HL earnings. (Total earnings from the BART already appeared on the computer screen during the BART activity.)

2.4 Empirical analyses

In our empirical analysis, we examine how well the different risk-taking propensity measures predict actual risk-taking behavior (smoking, drinking, seat belt non-use, and risky sex). To do this, we regress each risky behavior on a different risk-taking propensity measure (one at a time, each in a separate regression) while controlling

for sociodemographic variables (age, age squared, gender, education, marital status, income, numerical reasoning, importance of religion). We use a logistic regression model for actual behaviors, which have (0,1) dependent variables.

3 Results

Means of the sociodemographic variables are shown in Table 1, along with Spearman rank correlations with actual risk-taking behaviors. Thirty-seven percent of the respondents are male, with a mean age just under 32 years. Univariate statistics show that men are more likely to smoke and be problem drinkers, while older people are less likely to use a seat belt. Those for whom religion is more important are less likely to be problem drinkers. Education is also negatively correlated with smoking behavior.

Means and correlations between the actual risk-taking behaviors and the risk-taking propensity measures are shown in Table 2. Seventeen percent of respondents identified themselves as smokers, twenty-three percent were identified as problem drinkers using the AUDIT-C score, twenty-seven percent claimed to not use a seat belt the last time they sat in the front seat of a car, and twenty-eight percent were flagged as engaging in risky sexual behavior. Of note, except for drinking and smoking, which are highly correlated, the other combinations of health activities do not show high correlations.

3.1 Risk-taking propensity measures

Respondents, on average, reported themselves as 2.79 on a scale of 1 to 7 for the Dohmen measure. The distribution of Dohmen choices was skewed to the left, with far more people describing themselves as someone that tries to avoid risks than as someone prepared to take risks. This indicates greater risk aversion in our population than from the populations in Germany (Dohmen et al., 2011) or Senegal (Charness & Viceisza, 2011). Although men had higher Dohmen scores than women (3.02 vs. 2.65, respectively, not reported in the table), this difference was not statistically significant. The HL measure has a mean of 0.35, which is close to the estimates found in other studies (e.g., 0.3 to 0.5 in Holt and Laury (2002) in the U.S., 0.257 in Anderson and Mellor (2008) in the U.S., and 0.15 in Lammers (2008) in South Africa). The average BART earning was about R32. This is close to the (currency converted) amount earned by men in Lejuez et al. (2002), which was higher than that earned by women. In our sample, men and women scored the same on the BART. Nevertheless, because the earnings from an income maximizing fixed strategy of pumps would have

Table 1: Means of sociodemographic variables and correlations with actual risk-taking behaviors

Sociodemographics	Mean	S.D.	Correlations with actual risky behaviors			
			Smoking	Problem drinking	Seatbelt non-use	Risky sex
Male gender	0.37	0.48	0.28**	0.26**	-0.04	0.04
Age	31.61	9.92	0.10+	-0.09	-0.18**	0.06
Married/cohabiting	0.40	0.49	0.01	-0.03	-0.06	0.03
Numeracy	2.21	0.98	0.06	0.01	-0.05	-0.01
Religiosity	0.79	0.41	-0.06	-0.19**	-0.08	-0.02
Education (median)	(Some secondary)		-0.16**	0.01	0.10+	-0.06
Income per month (median)	(R2,000 to R4,000)		-0.09+	0.03	0.01	-0.03

+p<0.10, *p<0.05, **p<0.01, based on Spearman rank correlation; Sample size N=351.

earned R48, our respondents under-pumped, just as in Lejuez et al. (2002).³ The DOSPERT scores show relatively low willingness to take hypothetical risks, although participants were more willing to take non-health risks than health risks (score of 3.06 vs. 1.63; $t = 20$; $p < .01$).

Table 2 also presents the Spearman rank correlations between the various actual risk taking and risk-taking propensity measures in our data. The Dohmen measure was positively correlated with actual problem drinking and seat belt non-use. The HL and BART were not correlated at the 5% level with any of the actual risk-taking behaviors. The DOSPERT-*health* measure was correlated with actual smoking, problem drinking, and seat belt non-use, and the DOSPERT-*nonhealth* measure was not correlated with any of the actual risky health behaviors. The DOSPERT-*general* (which is a composite of both DOSPERT-*health* and -*nonhealth* scores) was correlated with smoking and drinking. The last row of Table 2 presents the correlations between the actual risky behaviors and the DOSPERT-*health* scores after deletion of the respective item in the DOSPERT that is similar to the risky behavior being correlated; for instance, when correlation with smoking is examined, the “DOSPERT-*health minus*” scores are constructed using the DOSPERT-*health* items except the smoking item. There are no significant correlations between actual risky behaviors and the DOSPERT-*health minus* measure.

Correlations between Dohmen, HL, and BART are not statistically significant at conventional levels, but Dohmen was fairly highly correlated with the DOSPERT measures. Some of the DOSPERT measures were corre-

lated with the HL at the 10% level.

3.2 Predictive power

The predictive power of each of the risk-taking propensity measures is shown in Table 3, derived from logistic regressions that controlled for all sociodemographic covariates. All models show the abilities of each of the measures in predicting self-reported actual health risk-taking behavior. The Dohmen measure is a statistically significant predictor of problem drinking and seat belt non-use, and almost predicts smoking and risky sexual behavior (with p-values of .068 and .065, respectively). The HL and BART measures are not significant predictors of any of the health behaviors. Although the DOSPERT-*general* and the DOSPERT-*nonhealth* measures do not predict actual risky behaviors, the DOSPERT-*health* measure is highly predictive of seat belt non-use and almost predicts smoking and problem drinking (with p-values of 0.088 and 0.052, respectively). However, as shown in Model 7, when we delete from the DOSPERT health domain an item (e.g., hypothetical seat belt non-use) that is similar to the dependent variable (e.g., actual seat belt non-use), the measure is no longer predictive of any of the actual risky health behaviors.

4 Discussion

Our results suggest that the Dohmen survey question on general risk-taking propensity is a good predictor of actual risky health behavior. Dohmen et al. (2011) found the general measure to be strongly related to investment in stocks, participation in active sports, self-employment, and smoking. This suggests that the measure can predict behaviors across different domains. Our findings fur-

³A fixed strategy of 64 pumps for each balloon maximizes earnings on the BART (Lejuez et al., 2002). With 30 balloons, 15 will explode before 64 pumps, and 15 will explode after 64 pumps. Therefore, only 15 balloons are banked: $(0.05 \text{ rand per pump}) \times (64 \text{ pumps per balloon}) \times (15 \text{ balloons banked}) = R48$.

Table 2: Means and correlations between actual risk taking and risk-taking propensity measures

	Mean	S.D.	Actual risk taking				Risk-taking propensity measures				
			1	2	3	4	5	6	7	8	9
Actual risk taking											
1. Smoking	0.17	0.37	—								
2. Problem drinking	0.23	0.42	0.35**	—							
3. Seat belt non-use	0.27	0.44	-0.05	0.01	—						
4. Risky sex	0.28	0.45	0.08	0.02	0.05	—					
Risk-taking propensity measures											
5. Dohmen	2.79	2.19	0.06	0.12*	0.15**	0.09+	—				
6. HL	0.35	0.62	-0.02	0.03	-0.00	-0.07	0.02	—			
7. BART	31.70	9.26	-0.01	0.10+	0.05	0.05	-0.04	-0.09+	—		
8. DOSPERT-general	2.26	0.79	0.12*	0.15**	0.07	0.03	0.32**	-0.11+	-0.02	—	
9. DOSPERT-health	1.63	0.73	0.14*	0.18**	0.13*	0.09	0.18**	-0.10+	-0.01	0.66**	—
10. DOSPERT-nonhealth	3.06	1.31	0.06	0.08	0.00	-0.01	0.29**	-0.08	-0.05	0.88**	0.27**
11. DOSPERT-health minus	—	—	0.09	0.09+	-0.00	0.08	—	—	—	—	—

+p<0.10, *p<0.05, **p<0.01, based on Spearman rank correlation.

Sample size N=351, except N=348 for BART and N=350 for DOSPERT-general and DOSPERT-nonhealth.

ther suggest that the Dohmen measure is able to predict a variety of different unhealthy activities *within* the health domain—despite variability in the types of behaviors and low correlations between them (except between drinking and smoking).

Our second finding is that the HL measure does not predict risky behavior in the domain of health, in some contrast to previous findings by Anderson and Mellor (2008). Despite their much bigger sample size (and thus greater power), the HL measure in their study was able to predict smoking, heavy drinking, or seat belt non-use only at the 10% significance level. Charness and Viceisza (2011) cast doubt on whether the participants in their sample in Senegal fully understood the HL task, as the distribution of choices was very different from that found in the U.S. population surveyed by Holt and Laury (2002). Because our study supervisors spent considerable time teaching the respondents about the HL task, with pictorial diagrams and actual props including two buckets and balls of different values, we are confident that most of our respondents understood the HL task. In fact, one of the pairs in the HL task contains a dominated option, and choosing this option is indicative that the respondent may not have understood the task. This occurred in 4.8% of our sample (17 out of 351), where the median member of our sample did not complete high school. In comparison, 4.2% of Anderson and Mellor’s sample (2008) chose the dominated option, even though over 70% of their respondents had some college education or beyond. In contrast, 17% of Lammers’ (2008) sample, who were

all college students in South Africa, chose the dominated option, and 40% of the Charness and Viceisza’s (2011) sample chose the dominated option, indicating that indeed the study participants in Senegal did not understand the HL task. Given the low percentage of respondents in our sample who chose a dominated option, we feel confident that non-understanding of the HL task is not the reason for lack of a relationship between the HL measure and health risk-taking behavior in our sample.⁴ For researchers who do use HL, we strongly encourage the use of detailed explanations with the aid of diagrams, instead of the usual tables that describe the trade-offs, which are much more difficult to understand. See Appendix Figures 1a and 1b for a comparison of the stimuli. If the HL questions are to be self-administered (as opposed to administered by an interviewer), instructions must be provided to the respondents before they view the diagrams. A detailed script that was used by the supervisors to administer the HL can be requested from the authors.

Our third finding is that the BART was not predictive of risky behavior in the domain of health in our context. This is in contrast to a large number of studies by Lejuez and co-authors (e.g., 2002). Although our study sample surely differed in exposure and experience with computerized tasks from other respondent groups commonly administered the BART, our respondents did not under-pump more than respondents in other studies. In fact, our

⁴Omitting the 17 individuals who made the dominated choice actually further reduces the predictive ability of the HL for all four risky activities.

Table 3: Regressions of actual behaviors on risk-taking propensity measures

Risk-taking propensity measures	Smoking	Problem drinking	Seatbelt non-use	Risky sex
Model 1: Dohmen	0.131+ [0.072]	0.133* [0.063]	0.126* [0.058]	0.104+ [0.056]
Model 2: HL	-0.063 [0.254]	0.187 [0.230]	0.091 [0.218]	-0.197 [0.201]
Model 3: BART	0.001 [0.017]	0.024 [0.016]	0.015 [0.015]	0.016 [0.014]
Model 4: DOSPERT-general	0.355+ [0.192]	0.234 [0.173]	0.178 [0.165]	0.126 [0.154]
Model 5: DOSPERT-health	0.334+ [0.196]	0.347+ [0.179]	0.524** [0.174]	0.237 [0.166]
Model 6: DOSPERT-nonhealth	0.163 [0.122]	0.049 [0.108]	-0.065 [0.101]	0.013 [0.094]
Model 7: DOSPERT-health minus	0.251 [0.191]	0.141 [0.181]	0.261 [0.181]	0.205 [0.149]

Numbers shown are coefficients [standard errors] from logistic models. All models include an intercept, age, age squared, gender, education, married/cohabitating, income, importance of religion, and numeracy. Significance levels are: + significant at 10%; * significant at 5%; ** significant at 1%. Sample size N=351, except N=348 for BART and N=350 for DOSPERT-general and DOSPERT-nonhealth.

lack of a finding is in congruence with a few other recent studies using BART that also did not find it to be predictive of smoking (Dean et al., 2011) or drinking (Klassen, 2010).

Our fourth finding is that the health domain version of the DOSPERT was predictive of unhealthy behaviors, but only when the instrument included a hypothetical activity similar to the specific unhealthy activity being predicted. This finding raises the question of whether risk behaviors are *activity-specific* beyond being *domain-specific*, and this finding needs to be explored further given mixed evidence in the literature. For instance, the health domain part of the DOSPERT-G used by Hanoch et al. (2006), which does not contain a smoking or exercise item (Johnson et al., 2004), is able to differentiate smokers (those that take risky health behaviors) from a comparison group consisting of gym members (“health seekers”), athletes, gamblers, and investors. Because the sampling approach in Hanoch et al. (2006) differs from that used in our study, the results are not directly comparable. In contrast, the DOSPERT used by Zuniga and Bouzas (2005) likely (although we are not sure, because we do not have the actual instrument) contains an alcohol use item in the instrument; it is unclear if deletion of this item would make the remaining health domain part of the instrument inef-

fective in predicting problem drinking behavior.

Overall, these findings suggest that a general risk-taking propensity measure elicited from a survey question may be a valid measure of risk-taking propensity for activities in different domains (as found by Dohmen et al., 2011) as well as for different activities within the health domain (as found by our study). A health domain specific measure, such as the DOSPERT-*health*, may also be a valid measure of risk-taking propensity for different unhealthy activities. However, a measure derived from the monetary/gambling specific domain may not be consistently applicable to explain behavior outside of the monetary/gambling domain, such as health, even if it is incentive compatible in the monetary/gambling domain. In other words, risk taking is domain specific. Using a measure derived from one domain may not apply to another domain. It is unclear why DOSPERT-*health* loses its predictive ability when an item similar to the one being predicted is deleted from the DOSPERT. This is unlikely due to a lower consistency of the measure with deletion of an item. The Cronbach α for our 5-item DOSPERT-*health* was 0.632, and deletion of the seat belt item results in a Cronbach α of 0.627—not an appreciable change—but lower than the 0.75 Cronbach α in the 10-item health domain in Weber et al. (2002).

4.1 Limitations

One limitation of our study was that only health behavior was examined. As risky health behavior across activities was not highly correlated, this mitigates the concern that the Dohmen measure was “lucky” in that it happened to be related to one (and thus all) health measures; instead it was able to predict these noncorrelated measures. Another limitation was that our instrument had a fixed order. Because the HL and BART tasks came first but the Dohmen measure followed the DOSPERT questions, we cannot eliminate the possibility that the DOSPERT questions influenced the Dohmen measure. That is, when the respondent answered the Dohmen, they could have been primed into thinking more about unhealthy behaviors because they had just answered hypothetical questions on the riskiness of such activities. Because questions on the actual health behaviors came at the end of the questionnaire, they should not influence responses to any of the risk-taking propensity measures. A final limitation of our study was that our sample was representative only of respondents who go for HIV testing, as opposed to a general population. Our results should be interpreted with the specific nature of our sample in mind. Although some of our respondents may have poor health, and illness may have affected their ability to perform the tasks or respond to questions, our numeracy instrument at least partially controlled for such effects.

We urge researchers to test further the usefulness of the Dohmen general measure and the robustness of the HL or BART measures, both in explaining health related risk-taking behavior and risk taking in other contexts. We also urge researchers to explore whether the predictive ability of the domain specific DOSPERT is compromised when the instrument no longer contains a hypothetical activity that is similar to the behavior being explained. Because a one-item survey question is so much easier to administer than the task-based HL and BART or the multi-item DOSPERT, there is definitely utility to the Dohmen measure, if it is consistently found in further studies to be a robust measure of risk-taking propensity.

References

- Anderson, L. & Mellor, J. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics*, 27, 1260–1274.
- Appelt, K., Milch, K., Handgraaf, M., & Weber, E. (2011). The decision making individual differences inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgment and Decision Making*, 6, 252–262.
- Bromiley, P. & Curley, S. P. 1992. Individual differences in risk taking. In J. F. Yates (Ed.), *Risk-taking Behavior* (pp. 87–132). Chichester: John Wiley & Sons Ltd.
- Bush, K., Kivlahan, D., McDonell, M., Fihn, S., & Bradley, K. (1998). The audit alcohol consumption questions (audit-c): an effective brief screening test for problem drinking. *Archives of Internal Medicine*, 158, 1789–1795.
- Camerer, C. & Hogarth, R. (1999). The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty*, 19, 7–42.
- Charness, G. & Vicejsza, A. (2011). *Comprehension and risk elicitation in the field: Evidence from rural Senegal*. (IFPRI Discussion Paper No. 01135).
- Dean, A., Sugar, C., Hellemann, G., & London, E. (2011). Is all risk bad? Young adult cigarette smokers fail to take adaptive risk in a laboratory decision-making test. *Psychopharmacology*, 215, 801–811.
- Deck, C., Lee, J., Reyes, J., & Rosen, C. (2010). Measuring risk aversion on multiple tasks: Can domain specific risk attitudes explain apparently inconsistent behavior? December 2010.
- Dislich, F., Zinkernagel, A., Ortner, T., & Schmitt, M. (2010). Convergence of direct, indirect, and objective risk-taking measures in gambling. *Zeitschrift für Psychologie/Journal of Psychology*, 218, 20–27.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9, 522–550.
- Figner, B., Mackinlay, R., Wilkening, F., & Weber, E. (2009). Affective and deliberative processes in risky choice: age differences in risk taking in the Columbia card task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 709.
- Gordon, M. (2007). *Evaluating the Balloon Analogue Risk Task (BART) as a Predictor of Risk Taking in Adolescent and Adult Male Drivers*. (Master’s thesis). The University of Waikato, Hamilton, New Zealand. Retrieved from <http://waikato.researchgateway.ac.nz/>.
- Hanoch, Y., Johnson, J., & Wilke, A. (2006). Domain specificity in experimental measures and participant recruitment an application to risk-taking behavior. *Psychological Science*, 17, 300–304.
- Holt, C. & Laury, S. (2002). Risk aversion and incentive effects. *The American Economic Review*, 92, 1644–1655.
- Johnson, J., Wilke, A., & Weber, E. (2004). Beyond a trait view of risk taking: A domain-specific scale measuring risk perceptions, expected benefits, and perceived-risk attitudes in German-speaking populations. *Polish Psychological Bulletin*, 35:153–172.

- Klassen, B. (2010). *Does disinhibition mediate alcohol use and risk taking?* (Master's thesis). Wayne State University, Detroit. Retrieved from <http://gradworks.umi.com/14/74/1474789.html>.
- Lammers, J. (2008). *HIV/AIDS, risk and intertemporal choice*. The Netherlands: Tilburg University.
- Lejuez, C., Read, J., Kahler, C., Richards, J., Ramsey, S., Stuart, G., Strong, D., & Brown, R. (2002). Evaluation of a behavioral measure of risk taking: the balloon analogue risk task (BART). *Journal of Experimental Psychology: Applied*, 8, 75.
- Reynolds, B., Ortengren, A., Richards J. B., & de Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and Individual Differences*, 40, 305.
- Weber, E. U. (2010). Risk attitude and preference. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 79–88.
- Weber, E. U., Blais, A., & Betz, N. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15, 263–290.
- Weber, E. U. & Johnson, E.J. (2008). Decisions under uncertainty: Psychological, economic, and neuroeconomic explanations of risk preference. In P. Glimcher, C. Camerer, E. Fehr, & R. Poldrack (Eds.), *Neuroeconomics: Decision making and the brain* (pp. 127–144). New York: Elsevier.
- Zuniga, A. & Bouzas, A. (2005). Actitud hacia el riesgo y consume de alcohol de los adolescented. *Manuscript submitted for publication*.

Appendix

Figure 1a: HL diagrams.

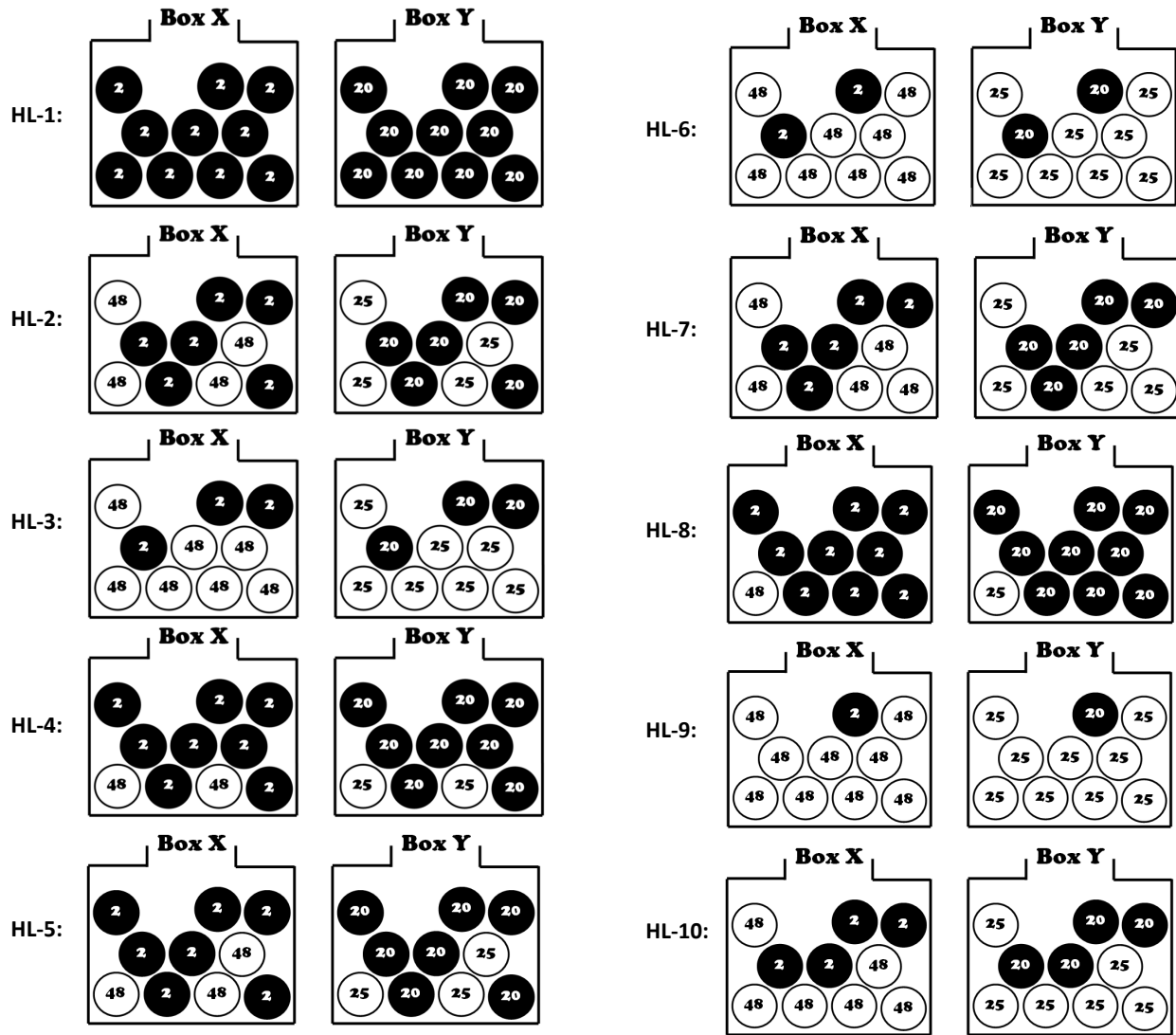
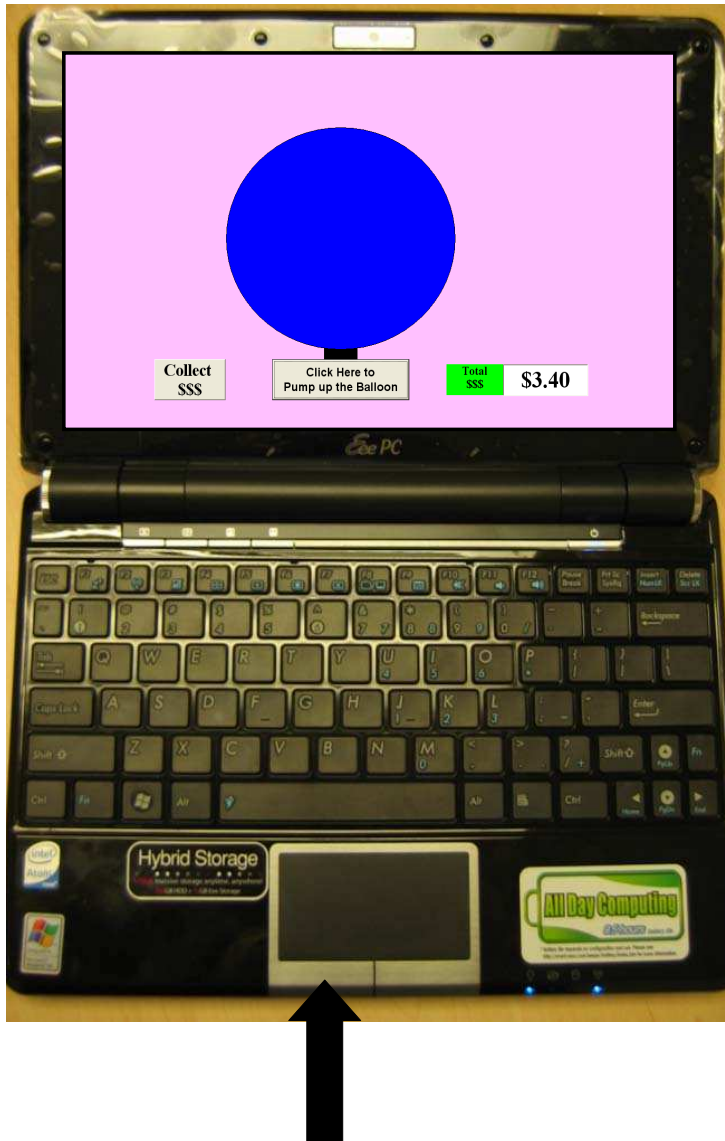


Figure 1b: Most common administration of HL task. Original Holt and Laury (2002), adapted to South African Rand.

Choice X	Choice Y
1/10 of R48, 9/10 of R2	1/10 of R25, 9/10 of R20
2/10 of R48, 8/10 of R2	2/10 of R25, 8/10 of R20
3/10 of R48, 7/10 of R2	3/10 of R25, 7/10 of R20
4/10 of R48, 6/10 of R2	4/10 of R25, 6/10 of R20
5/10 of R48, 5/10 of R2	5/10 of R25, 5/10 of R20
6/10 of R48, 4/10 of R2	6/10 of R25, 4/10 of R20
7/10 of R48, 3/10 of R2	7/10 of R25, 3/10 of R20
8/10 of R48, 2/10 of R2	8/10 of R25, 2/10 of R20
9/10 of R48, 1/10 of R2	9/10 of R25, 1/10 of R20
10/10 of R48, 0/10 of R2	10/10 of R25, 0/10 of R20

Figure 2a: BART diagram used to describe tasks.**Figure 2b: BART script used by interviewers.**

We have a new activity for you. You will use the computer for this activity, and you will get a chance to earn more money.

[POINT TO BALLOON DIAGRAM]

For this activity, you're going to see 30 balloons, one after another, on the screen. For each balloon, you will use the left button on the computer to pump up the balloon. Each click on the left button will pump the balloon up a little more.

But remember, balloons will explode if you pump them up too much. It is up to you to decide how much to pump up each balloon. Some of these balloons might explode after just one pump. Others might not explode until they fill the whole computer screen.

You get MONEY for every pump. Each pump earns 5 cents. But if a balloon explodes, you lose the money you earned on that balloon. To keep the money from a balloon, stop pumping before it explodes and click on the box labeled "Collect". This will bank the money you earned for that balloon.

After each time you collect or pop a balloon, a new balloon will appear.

At the end of the experiment, you will be paid the total amount earned on all the balloons you banked.

Figure 3: Numeracy questions.

-
0. Which is heavier: 100 kg of feathers or 100 kg of steel?
 1. Tomatoes at Checkers sell for 21 cents per kg. What will 4 kg of tomatoes cost?
 2. A boy is 6 years old and his sister is twice as old. When the boy is 10 years old, what will be the age of his sister?
 3. A patch of weed in a garden grows and doubles in size every day. If it takes the weed 48 days to cover the whole garden, how many days does it take the weed to cover half the garden?
 4. True or False: Two chickens and two dogs have a total of 14 legs.
 5. Thandiwe needs 13 bottles of water from the store. She can only carry 3 at a time. What is the minimum number of trips she needs to make to the store?
-

Note: Question 0 was asked to relax the respondent and was not part of the numeracy score. After answering the question, they were told the correct response ("the same").