NBER WORKING PAPER SERIES

RISK PREFERENCES OVER HEALTH: EMPIRICAL ESTIMATES AND IMPLICATIONS FOR HEALTHCARE DECISION-MAKING

Karen Mulligan
Drishti Baid
Jason N. Doctor
Charles E. Phelps
Darius N. Lakdawalla

Working Paper 31524
http://www.nber.org/papers/w31524

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2023

This work was supported by the Value of Life Sciences Innovation (VLSI) program at the Schaeffer Center for Health Policy and Economics at the University of Southern California. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w31524

Risk Preferences Over Health: Empirical Estimates and Implications for Healthcare Decision-Making
Karen Mulligan, Drishti Baid, Jason N. Doctor, Charles E. Phelps, and Darius N. Lakdawalla
NBER Working Paper No. 31524
August 2023
JEL No. I11,I18

ABSTRACT

Recent research has documented a link between consumer risk preferences over health and the willingness to pay (WTP) for medical technologies. However, the absence of empirical health risk preference estimates so far limits the implementation of this generalized risk-adjusted cost-effectiveness (GRACE) theory, which addresses several limitations of traditional cost-effectiveness analysis (CEA). To address this gap, we elicit from a nationally representative U.S. sample individual risk preference parameters over health-related quality of life (HRQoL) that shed light on health risk attitudes and enable GRACE valuation of medical technology. We find individuals exhibit risk-seeking preferences at low levels of health, switch to risk-averse preferences at health equal to 0.485 (measured on a zero to one scale), and become most risk-averse when their health is perfect (coefficient of relative risk aversion = 4.36). The risk preference estimates imply an empirical premium for disease severity: each unit of health is worth three times more to patients with serious health conditions (health equals 0.5) than those who are perfectly healthy. They also imply that traditional CEA overvalues treatments for the mildest diseases by more than a factor of two. Use of traditional CEA both overstimulates mild disease treatment innovation and underprovides severe disease treatment innovation.

Karen Mulligan
Leonard D. Schaeffer Center for
Health Policy & Economics
635 Downey Way
Verna & Peter Dauterive Hall (VPD)
Los Angeles, CA 90089
karenmul@usc.edu

Drishti Baid
Leonard D. Schaeffer Center for
Health Policy & Economics
635 Downey Way
Verna & Peter Dauterive Hall (VPD)
Los Angeles, CA 90089
dbaid@usc.edu

Jason N. Doctor
Schaeffer Center For Health Policy
and Economics
University of Southern California
635 Downey Way
Los Angeles, CA 90089-3333
jdoctor@usc.edu

Charles E. Phelps
University of Rochester
7 Millay Circle
Pittsford, NY 14534
charles.phelps@rochester.edu

Darius N. Lakdawalla
University of Southern California
635 Downey Way, VPD 414-K
Leonard D. Schaeffer Center for
Health Policy and Economics
Los Angeles, CA 90089-7273
and NBER
dlakdawa@usc.edu

Risk preferences over health: empirical estimates and implications for healthcare decision-making*

Karen Mulligan
Drishti Baid
Jason N. Doctor
Charles E. Phelps
Darius N. Lakdawalla

July 25, 2023

Abstract

Recent research has documented a link between consumer risk preferences over health and the willingness to pay (WTP) for medical technologies. However, the absence of empirical health risk preference estimates so far limits the implementation of this generalized risk-adjusted cost-effectiveness (GRACE) theory, which addresses several limitations of traditional cost-effectiveness analysis (CEA). To address this gap, we elicit from a nationally representative U.S. sample individual risk preference parameters over health-related quality of life (HRQoL) that shed light on health risk attitudes and enable GRACE valuation of medical technology. We find individuals exhibit risk-seeking preferences at low levels of health, switch to risk-averse preferences at health equal to 0.485 (measured on a zero to one scale), and become most risk-averse when their health is perfect (coefficient of relative risk aversion = 4.36). The risk preference estimates imply an empirical premium for disease severity: each unit of health is worth three times more to patients with serious health conditions (health equals 0.5) than those who are perfectly healthy. They also imply that traditional CEA overvalues treatments for the mildest diseases by more than a factor of two. Use of traditional CEA both overstimulates mild disease treatment innovation and underprovides severe disease treatment innovation.

1. INTRODUCTION

Orthodox cost-effectiveness analysis (CEA) implies that a "QALY is a QALY is a QALY…" [1]. That is, quality-adjusted life-years (QALYs) are equally valuable, regardless of the context in which they are added. Empirical evidence against this implication has accumulated. In population surveys, respondents express a preference for allocating QALYs towards more severe illness states [2, 3]. Studies of health technology assessment (HTA) and allocation regimes find a similar pattern of decision makers prioritizing health improvements for sicker groups [4, 5]. Revealed preference studies of severely ill patients with high cost-sharing plans suggest much higher willingness to pay than what is typically assumed by CEA [6]. Theorists and practitioners of cost-effectiveness have acknowledged these concerns too. Expert panels have suggested tempering QALY-based analysis by presenting results alongside

* The authors would like to thank Arthur Attema and Han Bleichrodt for their helpful comments and suggestions. All errors are our own.

evaluations of equity and justice or by conducting sensitivity analysis that allows for alternative valuations [7, 8]. Yet, even though HTA decisions often skew towards caring for the severely ill, methods for incorporating this feature tend to be qualitative and to depart from microeconomic foundations [9]. Meanwhile, quantitative approaches each tend to focus on addressing one or more specific limitations of the QALY; for example, the "value flower" enumerates these many deficiencies [10], but implementing the range of proposed "fixes" risks introducing double-counting or other errors into value assessment.

To help address these issues, Lakdawalla and Phelps have developed the theory of generalized risk-adjusted cost-effectiveness (GRACE) [11-14]. Traditional CEA often models consumers as if they are risk-neutral over health-related quality of life (HRQoL) [15].[1] GRACE relaxes this restriction and allows for arbitrary risk preferences, including risk-averse and risk-seeking behavior. Risk-aversion, and the attendant diminishing returns to health improvement, would explain decision makers' tendencies to place more value on improving health in sick states than traditional CEA implies. At the same time, the potential for risk-seeking behavior could reconcile theory with empirical evidence of the "value of hope" [17, 18], wherein some patients seem to prefer risky therapies with a chance of upside over risk-free alternatives with the same expected value.

So far, however, GRACE and the implied relationship between risk preferences and the value of health improvement has remained theoretical, because key parameters measuring risk preferences over HRQoL improvement have been unknown. In this paper, we provide the first estimates of risk preferences designed to parameterize the more general GRACE framework and to reveal the resulting variation in the value of health improvement and medical technology. We also develop the theoretical justification for estimating these risk preferences in an expo-power utility framework [19]. To demonstrate heterogeneity in preferences, we present individual-level estimates of risk preference

[1] In some cases, consumers are modeled as being risk-neutral over utility from HRQoL, but risk-averse over an underlying health index [16]. This leads to similar implications for the determinants of willingness to pay for HRQoL. As an aside, however, this approach tends to violate the implication that the risk premium on an uncertain health index varies with the probability of health losses, because HRQoL is not generally estimated as an arbitrary function of the probability of a poor health state.

parameters and test for differences in the estimated relative risk coefficient by demographics. Finally, we show how our estimates impact the willingness to pay (WTP) for marginal and inframarginal health improvements and discuss implications for CEA and HTA. The estimates in our paper, coupled with the formulae presented in prior theoretical research and the usual parameters needed for a traditional CEA [14], enable practitioners to implement GRACE analyses for any medical technology of interest.

While ours is the first study capable of operationalizing GRACE, it contributes to an established literature on empirical risk preferences over health. Health in these studies is often measured as longevity [20-23], but some studies have used alternative measures like HRQoL, QALYs, and pain [24-28]. Irrespective of the health measure used, individual preferences were generally consistent with nonlinear (i.e., non-risk-neutral) utility over health. Other studies have tested the validity of expected utility theory [20, 24, 26, 27] or explored the correlation between risk preferences over HRQoL and longevity [21, 28]. In the study most similar to ours, Attema et al (2016) estimate risk preferences over HRQoL for a representative sample from the Netherlands. Depending on the question framing, they found 36%-62% of responses were consistent with risk-averse preferences, 15%-39% were consistent with risk-seeking preferences, and 0-44% were consistent with risk-neutral preferences.[2] The authors' estimated utility parameters imply a median coefficient of relative risk aversion ranging from 0.1 to 0.39. While these estimates serve as a starting point for GRACE implementation, they require constant relative risk aversion (CRRA), which may not appeal to all practitioners or apply to all populations of interest.[3] It is nonetheless worth noting that when we adopt their CRRA utility assumption, our estimates are consistent with theirs.

Our expo-power approach allows relative risk-aversion to vary with health: pooled estimates for this coefficient range from -1.14 at very low levels of health (health level of 0.1, measured on a 0 to 1 scale) to 4.51 at "perfect" health (health equal to 1.0). More generally, we find individuals are risk-

[2] Questions varied the reference health value and whether individuals faced a health gain, loss, or mixed prospect relative to the reference health.

[3] The authors assumed a power utility function ($U(H) = H^{\alpha}$).

seeking (risk-averse) at levels of health below (above) 0.485. Our findings suggest a willingness to gamble on risky therapies among patients in worse health states, consistent with the "value of hope" found by prior studies [17, 18].[4] They also align with higher WTP for the treatment of moderately severe illness and for treatments that provide some chance of major upside for highly severe illness [29]. We observe heterogeneity in preferences at the individual level, with larger variance in individual risk preference parameters at higher levels of health. However, we do not find evidence that risk preferences are strongly influenced by demographics. Furthermore, we find limited support for relying on utility functions that limit risk preferences to CRRA: only 20% of respondents exhibit CRRA preferences over health when we use a utility function that allows for a range of risk preference structures. Finally, for illnesses that result in health greater than 0.79 (i.e., health loss less than 0.21), the risk-adjusted WTP threshold for marginal health improvements lies below the corresponding WTP threshold from traditional CEA. Therefore, traditional CEA over-values treatment for mild diseases with quality of life above this cutoff.

The remainder of our paper is structured as follows. Sections 2 and 3 present our experimental design and analytical methods. Section 4 presents the pooled and individual estimated utility and risk preference parameters. Section 5 uses the risk preference results to explore the value placed on marginal and inframarginal health improvements and the resulting implications for CEA. Section 6 provides a discussion and concludes.

2. METHODS

2.1. Experimental design

The experimental design consisted of two treatments. In both treatments, subjects were told to imagine health as a number ranging from 0 to 100, where 100 corresponds to perfect health and 0 to a health state just as bad as being dead. Respondents were also told to imagine being 40 years old with

[4] The prior empirical literature on "value of hope" focuses on life-extension rather than HRQoL improvements. Our finding is the first to our knowledge suggesting a qualitatively similar preference over HRQoL improvement.

usual health equal to 100, and to imagine further that their health deteriorates to some specified level X (where X is less than 100); here, X is the "reference" level of health.[5] Use of a reference point allows us to test whether risk preferences systematically vary across HRQoL gains and losses [30]. We set 20 as the minimum possible health value in the survey because values close to zero have been shown to cause extreme behavior and confusion between health and life expectancy [26, 31, 32].

Respondents were given thirteen different "scenarios," each consisting of a prospect and a reference health level. Certainty equivalents were elicited for each scenario. The first six scenarios ("common questions") were common across all respondents. They consisted of two different prospects, ((0.5,25;45) and (0.5,55;75)), each of which was presented for three different reference health values. Henceforth, the notation (0.5,X;Y) refers to a prospect involving a 50/50 gamble over health levels X and Y. The reference health values were selected such that each prospect was separately framed to each respondent as a gain, as a loss, and as a mixed outcome.

The second treatment followed the design of the reference-dependent treatment used by Attema et al (2016) [26]. This treatment consisted of a set of seven scenarios, and respondents were randomly assigned to the "gains" arm or the "losses" arm. The gains arm used a reference health of 20, and the losses arm used a reference health of 100. In sum, thirteen scenarios were presented, all of which are reported in Table 1 below. The next section describes how certainty equivalents were elicited for each scenario.

2.2.Subjects and procedure

The choice experiment was fielded within the Understanding America Study (UAS). The UAS is a nationally representative internet panel of approximately 9,500 respondents aged 18 or older and administered by the Center for Economic and Social Research at the University of Southern California.[33] Our survey was reviewed and approved by BRANY IRB (protocol #22-030-1044-198196). We ran a small pilot (N=115) to test whether the survey worked and generated sensible

[5] We added a reference value for age in response to respondent feedback from our pilot survey.

responses. For the full survey, N=1,144 subjects participated. For the second treatment, N=558 subjects were assigned to the gains arm, and N=586 to the losses arm. Subjects were paid $10 each for their participation.

Demographic characteristics for UAS participants are collected regularly and were merged with our survey questions. The survey began with questions related to perceptions of health; this section was designed to familiarize respondents with questions about health, before introducing the concept of risk. Specifically, respondents were first asked to rate their own health on a scale of 0 to 100. Next, they were asked to rate their expected health in ten years, the health of their spouse/significant other, and the health of an average person their age. Respondents were asked to rate five health values ranging from 25 to 75 on a scale ranging from "not very bad" to "extremely bad".

For each of the 13 health scenarios, certainty equivalents (CE) were elicited by asking respondents to choose between the risky health prospect and a series of certain outcomes. Participants were told to assume the reference health level and presented with the risky treatment option on the right side and a series of outcomes associated with the certain treatment option on the left side. The question format for each scenario is illustrated below for a risky treatment with equal chances of having health equal to 20 or 40.

	I prefer Treatment A	I prefer Treatment B	
Treatment A: Certain outcome = 21			Treatment B: 20 (50% chance) 40 (50% chance)
Treatment A: Certain outcome = 24			Treatment B: 20 (50% chance) 40 (50% chance)
Treatment A: Certain outcome = 27			Treatment B: 20 (50% chance) 40 (50% chance)
Treatment A: Certain outcome = 29			Treatment B: 20 (50% chance) 40 (50% chance)
Treatment A: Certain outcome = 32			Treatment B: 20 (50% chance) 40 (50% chance)
Treatment A: Certain outcome = 35			Treatment B: 20 (50% chance) 40 (50% chance)
Treatment A: Certain outcome = 38			Treatment B: 20 (50% chance) 40 (50% chance)

For each row in the table, participants chose between the certain option (Treatment A) or the risky option (Treatment B). The range of certain option choices spanned the range of the gamble. The survey forced participants to have one of three possible response patterns: 1) select Treatment A in all rows; 2) select Treatment B in all rows; or 3) switch from Treatment B to Treatment A at a single point. At most one switching point between Treatment B and Treatment A was allowed per question, which ensured the CE fell between the endpoints of the gamble.

2.3.Stimuli

CEs were elicited from 13 different scenarios for all respondents (Table 1). Prior studies have found that respondents understand the choice task more easily when gains are presented first [26, 34]. In accordance with this finding, we presented all respondents with the same gains question first: 50/50 gamble over health levels of 25 and 45, starting from a reference level of 20. For respondents in the gains arm of the second treatment, the remaining questions were asked at random. Respondents in the losses arm of the second treatment were presented the remaining common questions in random order, and then given the second treatment, which was accompanied by a different set of instructions to frame the loss prospect.

All health values were contained in the interval [20,100]. Respondents were told to imagine having a hypothetical level of health (which corresponded to the reference health in each question) and were asked to choose between two treatments. Both treatments would change their health for one year, after which time it would return to 100. The risky treatment option would change health to one of two possible levels, each with a probability of 50%.[6] The certain treatment option would change health to a given value with 100% probability.

The instructions in the common questions asked subjects to imagine that their health is usually 100 but has deteriorated to some hypothetical level (ranging from 20 to 85). Doctors have discovered the

[6] We considered alternate survey designs such as the one used by Holt and Laury (2002), which hold the outcomes fixed and vary the probabilities across prospects [35]. One study has shown that this format is more accurate for eliciting the shape of the probability weighting function than the utility function [36]. Moreover, changing probabilities introduces additional complexity in the question and may be better suited toward respondents with higher levels of mathematical skill [37].

cause of the health deterioration, and two treatments are available. For the gains (losses) questions, one treatment involved a certain health gain (loss). The other treatment involves risk, producing a larger gain (loss) with 50% probability and a smaller gain (loss) with 50% probability. For the two mixed prospect questions, the risky treatment provided a health gain with 50% probability and a health loss with 50% probability.[7] The gains arm of the second treatment presented the same instructions as the common gains questions.

For the losses arm of the second treatment, participants were asked to assume their health would deteriorate to 10 in the following year, unless they received treatment. Following Attema et al (2016), [26] this feature was necessary to prevent inaction from becoming the optimal decision.[8] After one year, the disease would disappear naturally, and health would return to 100. However, two treatments are available that would reduce the size of the health loss in the coming year. One treatment involved a sure loss. The risky treatment gave a small or no loss with 50% probability and a larger loss with 50% probability. The full survey and instructions are available online in the survey codebook (https://uasdata.usc.edu/index.php, UAS 462).

Table 1. List of health prospects

Question	Prospect	Reference health
1	(0.5,25;45)	20
2		35
3		85
4	(0.5,55;75)	45
5		65
6		75
7	(0.5,20;40)	Arm 1: 100 Arm 2: 20
8	(0.5,20;60)	
9	(0.5,20;100)	
10	(0.5,30;70)	
11	(0.5,50;90)	
12	(0.5,80;100)	
13	(0.5,60;80)	

Note: Prospects are defined as $(0.5, H_1; H_2)$, indicating 50% probabilities assigned to each of the health levels, H_1 and H_2, respectively.

[7] During pilot testing, we used "side effects" as a conceptual way to frame why a treatment could cause health losses. However, some respondents felt additional information was needed (i.e., what were the specific side effects) to answer the question. To avoid respondents imposing their own framing about potential side effects, we refrained from using that language as a conceptual framework.

[8] The drop in health described in the losses arm is only conditional on inaction and should not be considered as a possible reference point.

3. ANALYSIS

3.1. Exclusion criteria

We excluded N=397 individuals who for questions 7 through 13 (Table 1) either always selected the gamble, always selected the certain outcome, or did not respond.[9] CEs and thus risk preferences are not identified for any of these excluded individuals, who were more likely to fall in the bottom decile of total time spent on the 13 health gamble questions.[10] We present a comparison of demographics for the included and excluded samples in Table 2.

Table 2. Summary statistics

	Included sample	Excluded sample			
		Always picked the certain treatment	Always picked the risky treatment	Mixture of always certain or always risky	Missing values for all questions
Sample size	747	126	54	188	29
Male	0.420 (0.493)	0.309 (0.464)	0.388 (0.492)	0.319 (0.467)	0.344 (0.483)
Age (years)	52.15 (16.06)	53.41 (17.71)	56.40 (14.70)	47.09 (15.71)	45.20 (19.27)
Age group: <40 years	0.259 (0.438)	0.261 (0.441)	0.148 (0.358)	0.377 (0.486)	0.482 (0.508)
Age group: 40-54	0.270 (0.444)	0.230 (0.422)	0.277 (0.452)	0.292 (0.456)	0.275 (0.454)
Age group: 55-64	0.196 (0.397)	0.206 (0.406)	0.222 (0.419)	0.186 (0.390)	0.103 (0.309)
Age group: 65+	0.273 (0.445)	0.301 (0.460)	0.351 (0.482)	0.143 (0.351)	0.137 (0.350)
Married	0.574 (0.494)	0.492 (0.501)	0.574 (0.499)	0.473 (0.500)	0.482 (0.508)
College graduate	0.527 (0.499)	0.269 (0.445)	0.388 (0.492)	0.223 (0.417)	0.517 (0.508)
White, non-Hispanic	0.709 (0.454)	0.5 (0.501)	0.574 (0.499)	0.478 (0.500)	0.379 (0.493)
Working	0.566 (0.495)	0.547 (0.499)	0.518 (0.504)	0.542 (0.499)	0.655 (0.483)
Income: <60K	0.389 (0.487)	0.515 (0.501)	0.611 (0.492)	0.643 (0.480)	0.413 (0.501)

[9] Because questions 1 through 6 represent a mixture of gains, losses, and mixed prospects, individuals could have feasibly selected all certain outcomes or all gambles under prospect theory. As a result, response patterns for questions 1 through 6 were not used for the exclusion criteria.

[10] 24.6% of the excluded individuals fell in the bottom decile of total time spent on the 13 health questions, and 42.4% fell in the bottom two deciles. For comparison, only 2.8% and 9.1% of people included in our analysis fell in the bottom decile or bottom two deciles, respectively, of total time spent on the 13 health questions.

Income: 60K-99.9K	0.293 (0.455)	0.277 (0.449)	0.166 (0.376)	0.196 (0.398)	0.241 (0.435)
Income: 100K+	0.317 (0.465)	0.206 (0.406)	0.222 (0.419)	0.159 (0.367)	0.344 (0.483)
Insured	0.716 (0.451)	0.579 (0.495)	0.611 (0.492)	0.478 (0.500)	0.517 (0.508)
Region: Northwest	0.133 (0.340)	0.079 (0.271)	0.111 (0.317)	0.111 (0.315)	0.068 (0.257)
Region: Midwest	0.231 (0.422)	0.150 (0.359)	0.148 (0.358)	0.180 (0.385)	0.068 (0.257)
Region: South	0.269 (0.443)	0.301 (0.460)	0.259 (0.442)	0.265 (0.443)	0.379 (0.493)
Region: West	0.364 (0.481)	0.468 (0.500)	0.481 (0.504)	0.441 (0.497)	0.482 (0.508)
Self-rated health: 0-75	0.357 (0.479)	0.428 (0.496)	0.388 (0.492)	0.510 (0.501)	0.172 (0.384)
Self-rated health: 76-85	0.350 (0.477)	0.190 (0.394)	0.222 (0.419)	0.196 (0.398)	0.103 (0.309)
Self-rated health: 86-100	0.291 (0.454)	0.380 (0.487)	0.388 (0.492)	0.292 (0.456)	0.724 (0.454)

Notes: Standard deviations in parenthesis.

3.2.Estimating utility

CEs were calculated for each scenario as the midpoint of the certain outcomes between the two adjacent rows in which a respondent switched from preferring the risky treatment to preferring the treatment with a certain outcome. The table below illustrates using an example. In the table, the individual chooses the risky treatment when the certain outcome is 24, but the certain treatment when the certain outcome rises to 27. In this case, the CE would be estimated as $\frac{24+27}{2} = 25.5$. For individuals who selected the gamble in all choice sets for a particular question, the CE was the midpoint of the largest gamble endpoint and the largest certain outcome (i.e., the certain outcome in the last choice set). For example, if an individual had preferred Treatment B for all rows in the table below, the CE would be computed as $\frac{38+40}{2} = 39$. Similarly, for individuals who selected the certain outcome in all choice sets for a particular question, the CE was the midpoint of the smallest gamble endpoint and the smallest certain outcome (i.e., the certain outcome in the first choice set). Continuing with the example below, the CE in this case would be $\frac{20+21}{2} = 20.5$.

	I prefer Treatment A	I prefer Treatment B	
Treatment A: Certain outcome = 21		X	Treatment B: 20 (50% chance) 40 (50% chance)
Treatment A: Certain outcome = 24		X	Treatment B: 20 (50% chance) 40 (50% chance)
Treatment A: Certain outcome = 27	X		Treatment B: 20 (50% chance) 40 (50% chance)
Treatment A: Certain outcome = 29	X		Treatment B: 20 (50% chance) 40 (50% chance)
Treatment A: Certain outcome = 32	X		Treatment B: 20 (50% chance) 40 (50% chance)
Treatment A: Certain outcome = 35	X		Treatment B: 20 (50% chance) 40 (50% chance)
Treatment A: Certain outcome = 38	X		Treatment B: 20 (50% chance) 40 (50% chance)

Our analysis proceeds under the assumption of expected utility theory that reference health does not matter, and it pools data for all respondents and questions.[11] We estimated utility over health under two different parametric utility structures: expo-power [19] and constant relative risk-aversion (CRRA).[12] Expo-power nests increasing and decreasing relative risk-aversion, along with constant, increasing, and decreasing absolute risk-aversion. Expo-power also approaches constant relative risk-aversion as its two parameters approach zero. We also employ the commonly used CRRA form to facilitate comparison to prior literature [26].

Expo-power utility takes the form:

$$W^{EP}(q) = c - \exp\{-bq^a\} \tag{1a}$$

Since utility is equal for the prospect $(0.5, q_1; q_2)$ and its associated certainty equivalent (q_{CE}), we can write:

$$c - exp(-bq_{CE}^a) = 0.5\big(c - exp(-bq_1^a)\big) + 0.5\big(c - exp(-bq_2^a)\big) \tag{1b}$$

Solving (1b) for CE_i yields the expo-power estimating equation for each individual respondent i:

[11] We estimated EU models separately for the gains and losses arms and found no difference in utility parameters (see Table 4). Estimation of prospect theory models and formal hypothesis testing of whether behavior is better explained by EU or prospect theory are outside the scope of this paper.

[12] We also explored the use of Hyperbolic Absolute Risk-Aversion (HARA) utility. However, HARA models did not reliably converge, likely because of model misspecification, as we discuss in the appendix (Section 7.6).

$$q_{CE}^{EP} = \left(-\frac{1}{b}\ln\left[\frac{exp(-bq_1^a)+exp(-bq_2^a)}{2}\right]\right)^{\frac{1}{a}} + \epsilon_i^{EP} \quad (2)$$

Two utility parameters (*a,b*) are simultaneously estimated using nonlinear least squares, with the restriction that $ab \neq 0$. Just as with OLS, the error process, ϵ_i is assumed to be Gaussian in nonlinear least squares. The third parameter (*c*) is not uniquely identified through estimation; we set it to 1 to satisfy the conventional requirement that utility in the state of zero health is zero.[13] This is also consistent with our instruction to survey respondents that they should imagine the health state of zero being just as bad as death. Uniqueness results and a representation theorem for Equation (1a) is given in the Appendix (Section 7.1).

CRRA utility takes the form, $U^{CRRA}(X) = \frac{x^{1-\rho}}{1-\rho}$. Analogously, its NLLS estimating equation is:

$$q_{CE}^{CRRA} = \sqrt[1-\rho]{\frac{q_1^{1-\rho}+ q_2^{1-\rho}}{2}} + \epsilon_i^{CRRA} \quad (3)$$

Estimating risk preferences

GRACE can be implemented once the HRQoL utility function is recovered.[14] Before presenting implications for GRACE, we first calculate relative risk-aversion ($r_H^* = -\frac{W''(H)}{W'(H)}H$), which reveals the cost borne by consumers when treatment effects vary [11]. The appendix (Section 7.2) also presents estimates for relative prudence, $\pi_H^* = -\frac{W'''(H)}{W''(H)}H$, which reveals the taste for positive skewness in treatment effects [11].[15] We calculated empirical estimates for the relative risk parameters over the full range of health [0-1]. Mathematical equations as well as parameter estimates for r_H^* and π_H^* are provided in the Appendix (Section 7.2). Standard errors for the risk parameter estimates were clustered at the individual-level and bootstrapped using 1,000 replications.

[13] Rosen (1988) explained that in models of mortality risk-reduction, the level of utility affects behavior, because the marginal utility of life-extension depends on the utility level [38]. In the economics literature on cost-effectiveness, it is conventional to assume that health-related utility is zero when health is zero [15].

[14] Taylor-Series approximations can be used to estimate GRACE-based value assessments using relative risk-aversion and relative prudence, even without knowledge of the specific utility function.[11] In what follows here, however, we exploit estimates of the utility functions to provide exact valuations of health improvement rather than imperfect Taylor-Series approximations [14].

[15] The latter can be interpreted as the "value of hope," where patients value risky treatments with the chance of greater upside [17, 18].

4. RESULTS

4.1. Pooled estimates of risk preferences

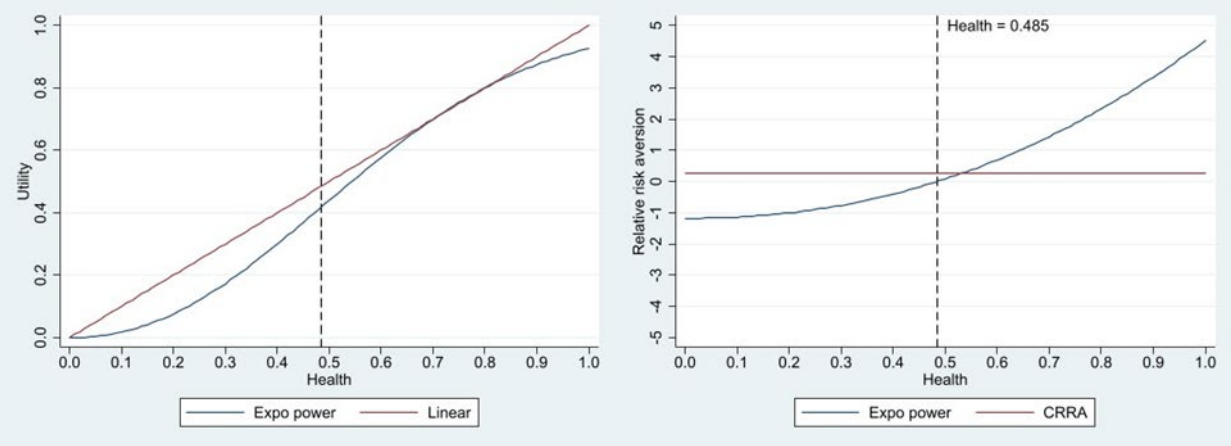
Table 3 presents results for the entire sample and shows the risk preference parameter estimates for select values of health (see Appendix Section 7.2 for variance-covariance matrices for utility estimates). Figure 1 presents the risk preference estimates for expo-power utility across the entire range [0,1] of health. The CRRA utility estimates yield a single value of $\boldsymbol{r}^*$ consistent with risk-aversion, while the expo-power utility estimates suggest individuals are risk-seeking (risk-averse) at low (high) levels of health. Under expo-power, preferences switch from risk-seeking to risk-averse at health equal to 0.485.

Table 3. Utility and risk parameter estimates, pooled sample

	Expo-power	CRRA
Utility parameters		
ρ	-	0.2822 [0.2170, 0.3492]
a	2.1760 [2.0687, 2.2883]	-
b	2.6152 [2.3936, 2.8609]	-
Relative risk aversion evaluated at select values for HRQoL(H)		
H = 0.1	-1.14 [-1.267, -1.005]	0.2822 [0.217, 0.349]
H = 0.5	-0.08 [-0.037, 0.198]	
H = 0.9	3.35 [2.846, 3.917]	
H = 1.0	4.51 [3.820, 5.332]	
Estimation N	9,710	9,710

Notes: Estimation sample pooled all 13 health gamble questions for N=747 respondents. 95% confidence intervals for the relative risk aversion parameters generated using 1,000 bootstrap replications clustered at the respondent level. Standards errors are similar without clustering. Utility parameter ρ corresponds to CRRA utility function: $U(H) = (H^{1-\rho})/(1-\rho)$; parameters a and b correspond to expo-power utility given by equation (1a).

Figure 1. Estimated utility and relative risk aversion parameters (pooled sample), full range of health [0,1]



Notes: Traditional CEA assumes utility that is linear in health-related quality of life, which is provided for reference as the red line in the left-hand panel. Relative risk aversion is calculated from utility parameter estimates for the expo-power and CRRA utility functions (Table 2). Expo-power preferences switch from risk-seeking to risk averse at a health value of 0.485.

We estimated expo-power utility parameters separately for the gains and losses arms for all health prospects as well for the subset of common questions and treatment questions (Table 4). We find no systematic differences in responses by treatment arm for common questions. Furthermore, we do not find evidence that framing a health prospect as again or a loss impacts the utility estimates; even the point estimates are quite similar across the gains and losses arms of the experiment.

Table 4. Expo-power utility parameter estimates by treatment arm

	All health prospects		Treatment questions		Common questions	
	Gains arm	Losses arm	Gains arm	Losses arm	Gains arm	Losses arm
a	2.217 [2.05, 2.38]	2.133 [1.97, 2.30]	2.292 [2.12, 2.46]	2.180 [2.01, 2.35]	1.948 [1.72, 2.18]	2.035 [1.81, 2.26]
b	2.721 [2.37, 3.08]	2.505 [2.16, 2.85]	2.837 [2.47, 3.20]	2.548 [2.19, 2.91]	3.014 [2.34, 3.69]	3.758 [3.01, 4.51]
N	5,004	4,706	2,694	2,534	2,310	2,172

Note: Estimation sample (N) pooled all 13 health gamble questions for 747 respondents.

4.2.Individual estimates of risk preferences

In principle, cost-effectiveness applies at the level of a representative individual. Inferences about insured populations rest on assumptions about homogeneity of individual preferences. Therefore, we quantify the heterogeneity in individual risk preferences and explore its possible determinants. Later, we

specifically analyze how well pooled risk preferences estimates approximate aggregate willingness to pay for health improvements.

Table 5 presents the summary statistics for the individual utility estimates. We first estimated individual expo-power utility functions for all respondents using all 13 health gamble questions pooled (results provided in Appendix Section 7.4). When either expo-power utility parameter (a or b) is close to zero, relative risk-aversion (r^*) is approximately constant over the full range of health (i.e., CRRA preferences).[16] Therefore, among the subsample of respondents (N=145) with at least one estimated expo-power parameter "close to zero" (defined as $|a| < 0.0001$ or $|b| < 0.0001$), we also estimated the CRRA utility function and used the utility function (expo-power or CRRA) with the smaller root mean squared error (RMSE) to calculate risk preference parameters for these individuals [39].

Table 5. Individual utility and relative risk aversion estimates, summary statistics

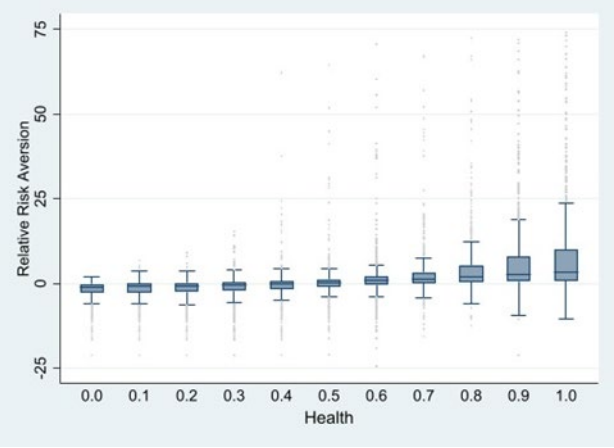
	Expo-power utility subsample		CRRA utility subsample	
	Median [IQR]	Mean (SD)	Median [IQR]	Mean (SD)
Utility parameters				
a	2.23 [1.39, 3.72]	3.65 (5.32)	1.39 [0.59, 2.71]	2.18 (3.21)
b	3.68 [1.74, 8.27]	12,338 (160,311)	3.5e-13 [3e-14, 2e-10]	4.1e-6 (1.6e-5)
ρ			-0.54 [-2.09, 0.25]	-1.15 (2.22)
Relative risk aversion evaluated at select values for HRQoL (H)				
H = 0.1	-1.18 [-2.72,-0.23]	-2.51 (5.41)	-0.54 [-2.09, 0.25]	-1.15 (2.22)
H = 0.5	0.40 [-0.94, 1.34]	1.82 (21.17)		
H = 0.9	4.44 [1.57, 14.1]	10,193 (118,188)		
H = 1.0	5.97 [2.09, 23.6]	448,111 (5,989,231)		
Respondents (N)	639		108	

Notes: First, the expo-power utility function was estimated for all individuals using all 13 health gamble questions. If either of the estimated utility parameters (a or b) was "close to zero" (defined as <0.0001 in absolute value), we estimated the CRRA utility model. The final model (CRRA or expo-power) for this subgroup (N=145) was selected using the smallest RMSE; CRRA utility minimized RMSE for 74% (N=108) respondents with either expo-power parameter close to zero.

[16] While the calculated r* values when a or b equal zero exist, marginal utility is undefined. As a result, expo-power utility can be represented as linear in health when a or b equal zero.

Figure 2 provides the distribution of individual estimates for risk aversion over the full range of health. At the individual level, parameters consistent with IARA-IRRA preferences are most common (72.6%), followed by DARA-CRRA (11.4%) and IARA-CRRA (8%). Among the 527 individuals who switch from risk seeking to risk averse, approximately 50% switch between a health level of 0.4 and 0.6. The distribution of individual estimates for r_H^* has higher variance at higher levels of health. Moreover, expo-power utility yields stronger degrees of relative risk aversion (i.e., larger magnitude of r_H^*) at higher levels of health compared with CRRA utility. For example, risk-averse CRRA individuals have a median $r_{H=1}^*$ equal to 0.52 [IQR=0.25, 0.81] compared with median expo-power $r_{H=1}^*$ equal to 7.18 [IQR=2.76, 27.0] for individuals with who are risk averse at perfect health ($H = 1$).

Figure 2. Individual relative risk aversion parameter estimates for full range of health [0-1]



Notes: Distribution of relative risk aversion parameters over the full range of health. The IQR is represented by the upper and lower edges of the rectangles. The whiskers represent the upper and lower adjacent values, which are calculated as the 75th percentile + 1.5*IQR and 25th percentile + 1.5*IQR, respectively. Individual risk preference parameters were derived from expo-power utility for the majority of respondents (N=639). The remaining N=108 respondents had expo-power utility parameter estimates that imply CRRA preferences and a smaller RMSE value for the CRRA utility model.

We estimated regressions to examine whether demographics were correlated with 1) absolute risk preference types; 2) relative risk preference types; and 3) relative risk aversion parameters. Regressions stratified by self-rated health subgroups suggest income and insurance status may be weakly correlated with preference type for people with self-rated health greater than 85 (which roughly corresponds to the top tercile of self-rated health in our sample). However, our results suggest relative risk aversion is relatively uniform across demographic groups (see Appendix Section 7.4 for results).

5. IMPLICATIONS FOR COST-EFFECTIVENESS ANALYSIS

5.1. Overview of GRACE concepts

The GRACE model extends CEA to explicitly and properly model non-risk-neutral preferences over health [11],[17] permanent disability [13], and ex ante decision-making [11]. While GRACE theory has been laid out elsewhere [11-14, 40], we provide a brief overview of GRACE here to explain how our empirical results relate to the conduct of cost-effectiveness analysis.

Under traditional CEA, the socially efficient WTP for HRQoL is the *ex ante* WTP for health improvement, which satisfies $K \equiv \frac{U(C)}{U'(C)H_0}$, where $U(C)$ is the utility of non-medical consumption, $U(C)$, and $H_0 = 1$ the *ex ante* HRQoL level [15]. In practice, healthcare decision makers may not have access to budgets that align with the socially efficient WTP threshold; in this case, K may lie below $\frac{U(C)}{U'(C)H_0}$ and be driven by budget constraints or other considerations. A technology is adopted if:

$$\frac{\Delta(Cost)}{\Delta(QALY)} \leq WTP \equiv K \tag{4}$$

Here, $\Delta(Cost)$ is the incremental cost of the technology, and $\Delta(QALY)$ is the incremental gain in quality-adjusted life-years. To define the QALY, denote HRQoL as H_T and H_U in the treated and untreated states, respectively, and denote the probability of survival as p_T and p_U in the treated and untreated states. The incremental QALY for a single period then satisfies: $\Delta(QALY) \equiv p_U(H_T - H_U) + H_T(p_T - p_U)$. For simplicity and without sacrificing generality, we present all our results for single-period QALY gains. Incremental QALYs gained over multiple periods can be calculated simply as the expected discounted sum of single-period gains.

By allowing for nonlinear utility over health, GRACE generalizes the above expressions. The generalized GRACE decision rule becomes:

[17] While some formulations of traditional cost-effectiveness accurately recognize the possibility of a nonlinear relationship between health and health-related utility [16], conventional cost-effectiveness analysis cannot be reconciled with basic predictions from expected utility theory. For instance, quality of life weights in conventional cost-effectiveness are assumed independent of the probability of a health loss and of the starting health level, even though both of these would obviously affect the risk premium associated with a given health loss.

$$\frac{\Delta(Cost)}{\Delta(GRA\text{-}QALY)} \leq K_{GRACE} \tag{5}$$

If the utility of health is given by $W(H)$, the generalized risk-adjusted QALY (GRA-QALY) for a single period can be written as: $\Delta(GRA\text{-}QALY) \equiv p_U \frac{(E(W(H_T))-E(W(H_U)))}{W'(H_U)} + \frac{E(W(H_T))}{W'(H_U)}(p_T - p_U)$ [14, equation 22c].[18] The first term in the GRA-QALY is the value of HRQoL improvement and the second the value of life-extension. Both allow for uncertainty in health states and associated risk-aversion. Note that the GRA-QALY reduces to the traditional QALY under linear utility [15], $W(H) = H$.[19]

Defining H_D as the time zero level of health, inclusive of any permanent disabilities or other pre-existing conditions, GRACE implies that the WTP for health improvements generalizes as [14, equation 22c]:

$$K_{GRACE} \equiv K \frac{W'(E(H_U))}{W(H_D)} H_0 \tag{6}$$

$\frac{W'(E(H_U))}{W(H_D)} H_0$ is the product of the elasticity of health-related utility with respect to health, $\omega_H \equiv \frac{W'(H_0)H_0}{W(H_0)}$, the "disease severity ratio," $R \equiv \frac{W'(E(H_U))}{W'(H_0)}$, and the "disability ratio," $D \equiv \frac{W(H_0)}{W(H_D)}$ [11]. The empirical estimate for ω_H using our pooled expo-power estimates is 0.4491.[20] For diseases of negligible severity ($R = 1$) afflicting consumers with no pre-existing disability ($D = 1$), WTP estimates should be more than halved, because $K_{GRACE} = K\omega_H \approx 0.4491K$. Therefore, GRACE implies that conventional cost-effectiveness overstates WTP whenever disease and/or pre-existing disability are mild enough that $RD < \frac{1}{\omega_H} \approx 2.227$. Later, we present estimates of R as a function of H_U, under the assumption of non-stochastic disease severity, and we present estimates of D as a function of H_D.

Expanding the expression for K_{GRACE} results in the equivalent reformulation:

---

[18] We provide readers with relevant equation numbers in Lakdawalla and Phelps (2023), abbreviated as LP, for reference. Because our paper focuses on the impact of risk preferences we assume disability (given by d* in LP (2023) equations), equals zero. Furthermore, we denote health in the treated state as H_T whereas LP denote it as $H_S + B$, which corresponds to health in the untreated state (H_U in this manuscript) plus the HRQoL gains (B) from treatment.

[19] The GRA-QALY also reduces to the traditional QALY if one defines K as the willingness to pay for marginal changes in health-related utility, as some traditional CEA theorists recommend [16].

[20] The empirical estimate for ω_H using the CRRA estimates is 0.7178.

$$K_{GRACE} \equiv K\omega_H RDH_0 \quad (7)$$

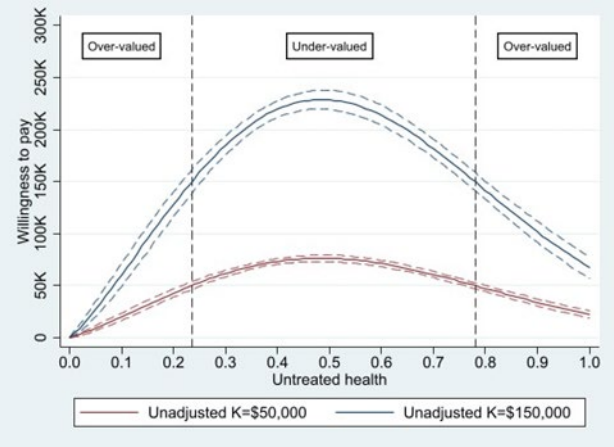
Evidently, K_{GRACE} is proportional to the conventional cost-effectiveness threshold, K. While the socially efficient level of K can be calculated using estimates for income and the elasticity of utility with respect to consumption from the literature, we present results in this section assuming K equal to \$50,000 and \$150,000, both of which are commonly used in the practice of CEA [41, 42].

5.2. Willingness to Pay for Health Improvements

Figure 3 presents the K_{GRACE} thresholds derived from the pooled expo-power utility estimates (Table 3); Appendix 7.5 presents a comparison of K_{GRACE} thresholds derived from expo-power and CRRA utility. These thresholds represent the value of marginal gains in health improvement; in practice, we use 0.01 unit improvements in HRQoL to measure "marginal" improvements. Since the WTP for health improvement varies with the marginal utility of health improvements, we expect it to rise in regions where people are risk-averse but fall where they are risk-seeking. Not surprisingly, therefore, the GRACE WTP threshold rises for HRQoL levels above 0.485 and falls for levels below it.

Moreover, for values of health greater than 0.78, the GRACE WTP threshold is less than the corresponding traditional CEA WTP threshold; thus, marginal gains for illnesses with untreated health between 0.78 and 1 will be overvalued under the risk-neutrality assumption of traditional CEA. In contrast, marginal gains for illnesses that cause untreated health to fall between 0.22 and 0.77 will be undervalued by traditional CEA. Finally, the downturn of WTP among the risk-seeking severely ill implies that CEA will overvalue marginal gains for illnesses with untreated health between 0 and 0.22. The dashed lines in Figure 3 represent the 95% confidence interval obtained from 1,000 bootstrap replications for the red and blue curves.

Figure 3. Generalized risk-adjusted WTP thresholds (pooled estimates)



Notes: 95% CIs for GRACE WTP thresholds derived from N=1000 bootstrap replications. For values of health less than 0.235 and health greater than 0.78, the adjusted WTP threshold is less than the corresponding unadjusted WTP threshold. This implies that CEA will overvalue health technologies that treat illnesses that correspond to a health of 0 to 23.5 and 78 to 100 and undervalue health technologies that treat illnesses that correspond to health of 24 to 77. In other words, using traditional CEA WTP thresholds, payers that rely on CEA would over-reimburse for treatments for relatively mild and extremely severe illnesses. For people with moderate to severe illness, payers would under-reimburse for treatments.

Table 6 presents the estimates for the disease severity ratio (*R*) and disability ratio (*D*) calculated directly from the pooled expo-power utility function estimates. If people were risk averse over the full range of health, we would expect *R* to increase as untreated health decreases. However, because people are risk seeking at low levels of health and risk averse at high levels of health, *R* has an inverted u-shape.

Table 6. Disease severity and disability ratios for selected values of HRQoL

Untreated health (H_U)	Disease severity ratio (R)	Untreated health, inclusive of disability (H_D)	Disability ratio (D)
1.0	1.00	1.0	1.00
0.9	1.51	0.9	1.06
0.75	2.41	0.75	1.23
0.5	3.39	0.5	2.11
0.25	2.36	0.25	7.71
0.1	0.90	0.1	53.61

Notes: Values are calculated using the pooled expo-power estimates (Table 3). The disease severity ratio is given by $\frac{W'(E(H_U))}{W'(H_0)}$ and the disability ratio is given by $\frac{W(H_0)}{W(H_D)}$. The calculations here assume $H_0 = 1$ and H_U is nonrandom.

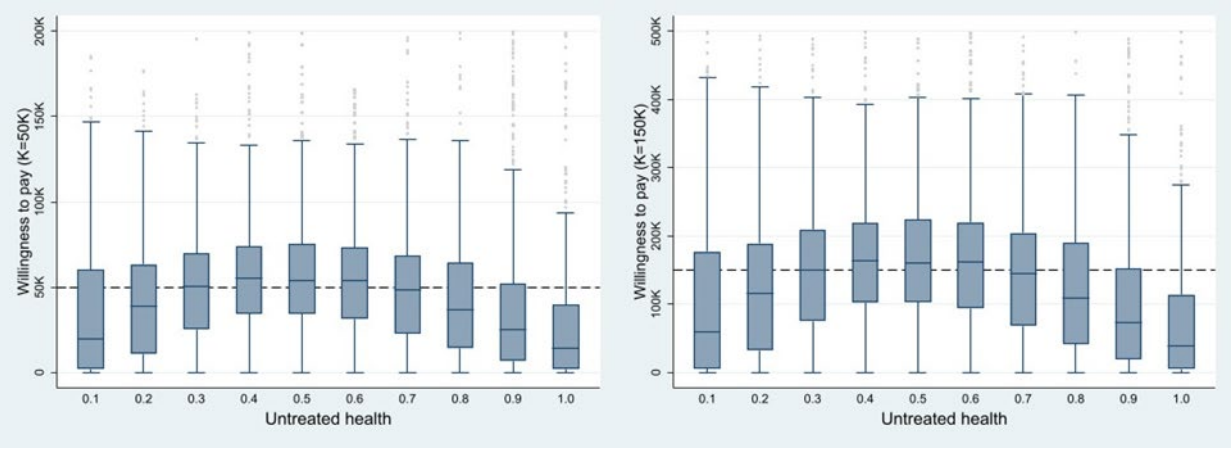
Table 6 demonstrates that treating illnesses at a health level of 0.5 generate more than triple the value per unit of QoL improvement than treating very mild disease. This contrasts starkly with traditional CEA, which implies no differences in the value per unit of QoL improvement. Similarly, the value of

health improvement rises with pre-existing disability. Even moderate disability of around H_D =0.75 increases the value of QoL improvement by nearly 25%. At the same time, the results illustrate the extent to which traditional CEA overvalues treatments for mild disease. Recall that $K_{GRACE} = K\omega_H RD$, and recall further that $\omega_H = 0.4491$ under expo-power utility. This implies that traditional CEA overvalues treatment of negligibly severe illness ($H_U \approx 1$) by a factor of 2.227. It overvalues mild illness ($H_U = 0.9$) by almost 50% ($\frac{1}{1.51*0.4491} \approx 1.475$). The disability adjustment also has meaningful implications for value. For instance, pre-existing peanut allergies ($H_D = 0.803$) would increase the value of all health improvements by 15% [43].

Below $H_U = 0.5$, Table 6 shows that the disease severity ratio begins to fall. This is due to risk-seeking behavior, which depresses $W'(E(H_U))$ and thus the severity ratio, R. Patients in these highly severe health states place less value on marginal improvements in health, but more value on marginal increases in the probability of large gains. We pursue the meaning of this distinction later.

Figure 4 shows the distribution of WTP thresholds based on the individual estimates. For the most part, the interquartile variation in WTP thresholds is relatively modest. For example, for a health level equal to 0.5, the IQR for the generalized adjusted WTP threshold is [$32,892, $76,180] and [$98,687, $228,541] corresponding to unadjusted K of $50,000 and $150,000, respectively. Similarly, at a health level equal to 1.0, the generalized adjusted WTP threshold IQR is [$669, $41,978] and [$2,008, $125,934].

Figure 4. Distribution of risk adjusted WTP thresholds (individual estimates)



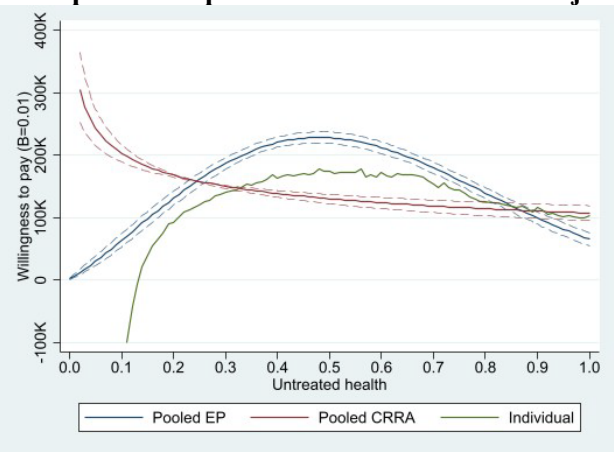
Notes: Distribution of risk adjusted WTP thresholds across range of health [0.1,1]. The IQR is represented by the upper and lower edges of the rectangles. The whiskers represent the upper and lower adjacent values, which are calculated as the 75th percentile + 1.5*IQR and 25th percentile + 1.5*IQR, respectively. Individual WTP estimates were derived from expo-power utility for the majority of respondents (N=639). The remaining N=108 respondents had expo-power utility parameter estimates that imply CRRA preferences and a smaller RMSE value for the CRRA utility model. Adjusted WTP threshold is undefined at health = 0. Y-axis scales for left and right panels are different, and have been truncated.

While consumers vary in their risk preferences, cost-effectiveness practitioners will rarely have access to individual-specific risk preference data in every population of interest. In such cases, the pooled expo-power estimates appear to be reasonable approximations to the aggregate WTP implied by the heterogeneous individual risk preferences across individuals. Figure 5 presents a comparison of the risk adjusted WTP based on the pooled estimates (expo-power and CRRA utility) and the aggregate WTP based upon the individual estimates. The individual risk-adjusted WTP reflects the average WTP across individuals; conceptually, this represents the aggregate willingness to pay for health improvement within the population, expressed in per capita terms by dividing by the total number of respondents (N=747). Since the confidence intervals around the green curve are so large as to extend beyond the axes in the figure, this analysis should be thought of as illustrative rather than formal.

The general shape of the individual WTP curve more closely resembles the WTP generated by pooled expo-power utility, which is consistent with the fact that approximately 86% of individuals in our sample had either non-CRRA preferences or CRRA preferences but a smaller RMSE for expo power utility. However, the figure also suggests the potential role of sensitivity analysis using both the expo-

power and the CRRA estimates to provide bounds on the individual estimates over most of the salient HRQoL domain.

Figure 5. Comparison of pooled and individual risk adjusted WTP



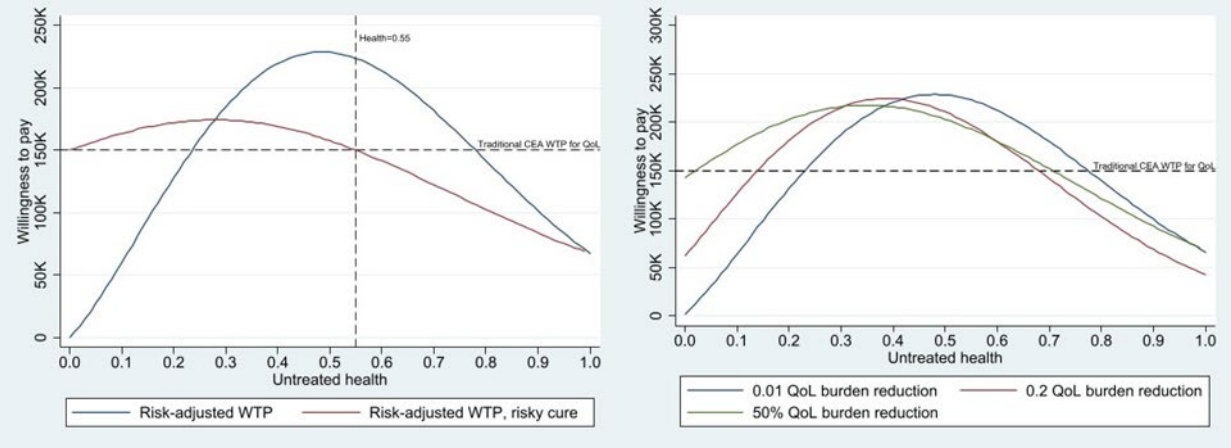
Notes: Individual curve corresponds to the per capita risk adjusted WTP derived from the individual level utility estimates summarized in Table 5. Dashed lines correspond to the 95% confidence intervals (CI); CIs for the individual WTP are wider than the y-axis scale and therefore omitted; we present a figure with the individual CIs included in the Appendix Section 7.5.

The results above report the value of marginal gains in HRQoL. However, the nonlinearity of the utility function suggests that marginal gains might not be worth the same as inframarginal gains. Meanwhile, the presence of variable risk preferences suggests differences between the value of marginal certain gains and of actuarially equivalent risky gains. Figure 6 explores these two hypotheses, and analogous comparisons of pooled and individual utility are presented in the Appendix Section 7.4.

The righthand panel compares (1) the willingness to pay for marginal gains (i.e., 0.01 unit gains in HRQoL), given by the blue curve, (2) the willingness to pay for inframarginal gains (defined here as 0.2 units of HRQoL), given by the red curve, and (3) the willingness to pay for 50% reductions in the HRQoL burden (i.e., HRQoL improvements of $\frac{H_0-H_U}{2}$), given by the green curve. All curves report the WTP per unit of HRQoL gain, so they are comparable in scale. Heuristically, "breakthroughs" are more valuable on a unit basis for the severely ill, because of their apparent risk-seeking posture. That is, inframarginal gains are relatively more valuable on a per unit basis than marginal ones among the severely ill. The opposite is true among those with mild disease.

The lefthand panel of Figure 6 illustrates how the value of risky gains departs from that of certain gains. The blue curve reproduces the value of a certain 0.01 unit HRQoL improvement. The red curve illustrates the WTP for a risky cure with an expected HRQoL gain of 0.01 units. In other words, the red and blue curves compare actuarially equivalent health improvements, but the red curve depicts a treatment with a small chance of curing the patient's condition entirely. Among the severely ill, the risky cure is relatively more valuable than the certain gain. This comports with earlier empirical research on the "value of hope" [17, 18].

Figure 6. Impact of risk and burden reduction on GRACE WTP thresholds



Notes: All WTP thresholds calculated assuming K=$150,000. Values are normalized by the size of the HRQoL gain from treatment. Left panel: Risk-adjusted WTP is the same as the blue line in Figure 3. The risky cure is a treatment with an x% chance of a cure, where $x = \frac{0.01}{1-H_U}$, which implies the expected HRQoL gain from treatment equals 0.01. Right panel: The GRACE WTP for a 0.01 HRQoL burden reduction approximates the risk-adjusted WTP (left panel). The 50% QoL burden reduction eliminates 50% of the QoL loss from the disease, while the 0.2 QoL burden reduction removes 0.2 units of loss.

6. DISCUSSION

Our study presents empirical estimates for risk preference parameters over health in a nationally representative U.S. sample. The pooled estimates indicate relative risk aversion is increasing in health. More specifically, individuals in the worst health state exhibit risk seeking preferences ($r^*_{0.1}$ equals -1.01), switch to risk-averse preferences at health equal to 0.485, and reach their maximum risk-aversion when their health is perfect ($r^*_{1.0}$ equals 4.36). Although there is a substantial degree of individual heterogeneity

in risk preferences over health, we find minimal evidence that risk preferences are correlated with common demographic covariates. In contrast, existing literature finds risk-aversion over consumption is correlated with gender and race [35, 44-46]. Our estimates have important implications for CEA and HTA, and suggest that traditional CEA will undervalue treatments for more severe illnesses and overvalue treatments for less severe ones.

To our knowledge, Attema et al (2016) provides the only published estimate of relative risk aversion (range: 0.1 to 0.39) over HRQoL [26]. Our r_H^* under CRRA utility is most directly comparable to their results, and our estimate of 0.28 lies near the middle of their range. While the consistency of our estimates suggests that risk preferences over health under the assumption of CRRA utility may be similar across different geographic settings and populations, additional studies are needed to further validate this finding. Furthermore, it may not be the case that relative risk preferences over health are non-constant in settings outside the U.S. Understanding the correct preference structure as well as empirical estimates for risk preference parameters will be important for obtaining accurate GRACE estimates to guide decision-making.

While GRACE addresses a key shortcoming of traditional CEA, it introduces additional elements for consideration when developing "best practices" for applied settings. We have highlighted the potential influence of underlying utility function used to derive risk preference parameters. There are other empirical decisions that mirror issues raised by the 2nd Panel on Cost-Effectiveness [47]. For example, the 2nd Panel recommended community preferences for health states are most appropriate for CEA analysis. Furthermore, generic preference-based measures such as EQ-5D are recommended to enhance comparability across studies. Analogous concerns are introduced for risk preference parameters in GRACE. Similar considerations must be made for the incorporation of risk preference parameters (community vs. patient) and how risk is measured. In our analysis, risk preferences were elicited based on a health scale ranging from 0 to 100, which provides a standardized measure. However, whether this measure adequately reflects risk preferences related to specific health attributes (e.g., poor vision, pain, or

mobility) is unknown. Additional studies that elicit risk preferences based on alternative measures for health will provide better insight into this issue.

Health levels below the switching point (H=0.485) for risk-seeking and risk averse preferences correspond to quite severe illnesses such as Alzheimer's disease, diabetes with end-stage renal disease or major amputation, Huntington's disease, and metastatic lung or colorectal cancer [48]. This result is consistent with motivations for patients who participate in Phase I clinical trials [49-51] as well as the premise of the Right to Try Act, which allows patients diagnosed with life-threatening illnesses who have tried all approved treatments to access investigational treatments [52]. Conversely, high degrees of risk aversion at relatively high levels of health could manifest in relatively low tolerance for side effects. Statin therapy for primary prevention provides a salient example,[53] since the measurable health loss (i.e., symptoms) associated with high cholesterol is negligible except in the most severe cases.

The medical field continues to shift toward shared decision making between providers and patients. The absence of evidence for risk neutrality over health implies patients would benefit from better understanding of uncertainty and risk associated with medical interventions. For example, a non-risk-neutral patient would prefer to make decisions based on the distribution of possible outcomes rather than median/average ones. A subtype of interventions, such as adjuvant therapy in cancer,[54] primarily serves to modify future risk and require patients to weigh risk structure (rather than a concrete HRQoL improvement) against side effects. Healthcare teams typically do not include personnel whose training is more directly focused on risk in medicine. One exception is genetic counselors, although their current scope is limited to areas that rely on genetic testing such as oncology and pediatrics [55, 56]. Introducing new providers who specialize in helping patients understand risk may improve patients' healthcare decisions and reduce the burden on physicians.

The range of individual heterogeneity in risk preferences suggests patients might benefit from a "top-up" insurance design that provides coverage for the cost of baseline treatments (selected based on some standard WTP threshold) and allows patients the option to pay the incremental cost of more expensive treatments[57, 58]. For example, assume baseline treatments are selected based on a WTP

threshold of $50K, which is currently used for coverage decisions in the U.K. For illnesses that result in health falling to 0.5, 58.6% of our sample has a risk-adjusted WTP greater than 50K and would likely select a "top-up" insurance design; because approximately 86% of our sample has expo power utility and exhibits risk seeking behavior at low levels of health, the share that would select a "top-up" design decreases to 38% for illnesses that cause health to fall to 0.2.

That said, top-up insurance models for health interventions with very low social marginal costs of production – like small-molecule prescription drugs -- may lower efficiency, compared to full and generous insurance coverage. In these contexts, exposing patients to on-patent prices leads to underutilization, and welfare can be improved with greater insurance [59]. Here, individual heterogeneity might be best addressed by providing more choice in the scope of coverage, where some policies feature coverage of technologies with higher costs per unit of health improvement and vice-versa.

Currently, HTA organizations evaluate efficiency and equity separately and present equity considerations in a qualitative manner alongside CEA results [60, 61]. Policy makers then determine the importance of elements related to equity such as disease severity or rarity through deliberative processes. An alternative approach involves incorporating equity preferences explicitly in CEA. Solutions that explicitly modify CEA include equity-based weighting, extended CEA, and distributional CEA; multi-criteria decision analysis (MCDA) and mathematical programming incorporate equity in a separate step from CEA analysis [62, 63]. Some may prefer these solutions to a deliberative approach because of their quantitative nature and the absence of discretion by third-party payers that may have financial incentives to limit coverage. However, each method still requires additional decisions such as which domains of equity to include, how they interact, and the relative weights given to equity and efficiency outcomes [64, 65]. In contrast, GRACE incorporates important health equity considerations without requiring these potentially controversial and normative decisions. Even if decisionmakers continue to rely on deliberative processes to consider elements beyond efficiency, GRACE could be used in lieu of CEA in the economic component of HTAs to better reflect patient preferences for risk, and the attendant implications for disease severity and pre-existing disability.

Our results were derived under the assumption that individuals behave according to expected utility theory. In contrast, prospect theory represents risky preferences as risk averse for gains and risk seeking for losses. In addition, there are also separate inverse-S shaped probability functions for gains and losses. Prospect theory would interpret this switch from risk-aversion to risk-seeking as a switch from coding gambles as gains to losses. While we found no difference in the estimated utility parameters across the gains and losses arms, we did not include controls for probability weighting. As a result, our finding cannot be interpreted as a rejection of prospect theory.

Our results indicate that CEA's reliance on risk neutral preferences does not align with real world preferences for risk over health. More specifically, a majority of respondents in our sample exhibited IRRA preferences over health and have a risk preference pattern that switches from risk seeking to risk averse at some point in the health distribution. From a policy perspective, this implies healthcare decisions based on traditional CEA have undervalued interventions for relatively severe illnesses and overvalues interventions for mild disease. The utility and relative risk aversion parameter estimates provided in this paper can serve as baseline estimates to implement GRACE, which provides a path forward to mitigate the misallocation of resources across disease severity that has afflicted cost-effectiveness analyses in the past.

REFERENCES

1. Williams, A., *Cost-Effectiveness Analysis: Is It Ethical?* Journal of medical ethics, 1992. **18**(1): p. 7-11.
2. Nord, E., *The Relevance of Health State after Treatment in Prioritising between Different Patients.* Journal of medical ethics, 1993. **19**(1): p. 37-42.
3. Nord, E., et al., *Maximizing Health Benefits Vs Egalitarianism: An Australian Survey of Health Issues.* Social Science & Medicine, 1995. **41**(10): p. 1429-37.
4. Nord, E. and R. Johansen, *Concerns for Severity in Priority Setting in Health Care: A Review of Trade-Off Data in Preference Studies and Implications for Societal Willingness to Pay for a Qaly.* Health Policy, 2014. **116**(2-3): p. 281-288.
5. Nord, E. and R. Johansen, *Transforming Eq-5d Utilities for Use in Cost–Value Analysis of Health Programs.* The European Journal of Health Economics, 2015. **16**: p. 313-328.
6. Seabury, S.A., et al., *Patients Value Metastatic Cancer Therapy More Highly Than Is Typically Shown through Traditional Estimates.* Health Affairs, 2012. **31**(4): p. 691-9.
7. Drummond, M., et al., *Toward a Consensus on the Qaly.* Value Health, 2009. **12 Suppl 1**: p. S31-5.
8. Brock, D.W., et al., *Ethical and Distributive Considerations*, in *Cost-Effectiveness in Health and Medicine*. 2016, Oxford University Press. p. 289-319.
9. Skedgel, C., et al., *Considering Severity in Health Technology Assessment: Can We Do Better?* Value in Health, 2022. **25**(8): p. 1399-1403.
10. Lakdawalla, D.N., et al., *Defining Elements of Value in Health Care-a Health Economics Approach: An Ispor Special Task Force Report [3].* Value Health, 2018. **21**(2): p. 131-139.
11. Lakdawalla, D.N. and C.E. Phelps, *Health Technology Assessment with Risk Aversion in Health.* Journal of Health Economics, 2020. **72**.
12. Lakdawalla, D.N. and C.E. Phelps, *Health Technology Assessment with Diminishing Returns to Health: The Generalized Risk-Adjusted Cost Effectiveness (Grace) Approach.* Value in Health, 2021. **24**(2): p. 244-249.
13. Lakdawalla, D.N. and C.E. Phelps, *A Guide to Extending and Implementing Generalized Risk-Adjusted Cost-Effectiveness (Grace).* Eur J Health Econ, 2022. **23**(3): p. 433-451.
14. Lakdawalla, D.N. and C.E. Phelps, *The Generalized Risk-Adjusted Cost-Effectiveness (Grace) Model for Measuring the Value of Gains in Health: An Exact Formulation.* Journal of Benefit-Cost Analysis, 2023. **In Press**.
15. Garber, A.M. and C.E. Phelps, *Economic Foundations of Cost-Effectiveness Analysis.* Journal of Health Economics, 1997. **16**(1): p. 1-31.
16. Bleichrodt, H. and J. Quiggin, *Life-Cycle Preferences over Consumption and Health: When Is Cost-Effectiveness Analysis Equivalent to Cost-Benefit Analysis?* Journal of Health Economics, 1999. **18**(6): p. 681-708.
17. Lakdawalla, D.N., et al., *How Cancer Patients Value Hope and the Implications for Cost-Effectiveness Assessments of High-Cost Cancer Therapies.* Health Affairs, 2012. **31**(4): p. 676-82.
18. Shafrin, J., et al., *Patient Versus Physician Valuation of Durable Survival Gains: Implications for Value Framework Assessments.* Value Health, 2017. **20**(2): p. 217-223.
19. Saha, A., *Expo-Power Utility: A 'Flexible'form for Absolute and Relative Risk Aversion.* American Journal of Agricultural Economics, 1993. **75**(4): p. 905-913.
20. Attema, A.E., B.F.B. Werner, and l.H. Olivier, *Prospect Theory in the Health Domain: A Quantitative Assessment.* Journal of Health Economics, 2013. **32**(6): p. 1057-1065.
21. Attema, A.E., O. l'Haridon, and G. van de Kuilen, *Measuring Multivariate Risk Preferences in the Health Domain.* J Health Econ, 2019. **64**: p. 15-24.

22. Delprat, G., M.L. Leroux, and P.C. Michaud, *Evidence on Individual Preferences for Longevity Risk.* Journal of Pension Economics & Finance, 2016. **15**(2): p. 160-179.
23. Galizzi, M.M., M. Miraldo, and C. Stavropoulou, *In Sickness but Not in Wealth: Field Evidence on Patients' Risk Preferences in Financial and Health Domains.* Med Decis Making, 2016. **36**(4): p. 503-17.
24. Schosser, S., J.N. Trarbach, and B. Vogt, *How Does the Perception of Pain Determine the Selection between Different Treatments?: Experimental Evidence for Convex Utility Functions over Pain Duration and Concave Utility Functions over Pain Intensity.* Journal of Economic Behavior & Organization, 2016. **131**: p. 174-182.
25. Prosser, L.A. and E. Wittenberg, *Do Risk Attitudes Differ across Domains and Respondent Types?* Med Decis Making, 2007. **27**(3): p. 281-7.
26. Attema, A.E., et al., *An Elicitation of Utility for Quality of Life under Prospect Theory.* Journal of Health Economics, 2016. **48**: p. 121-134.
27. Rouyard, T., et al., *Risk Attitudes of People with 'Manageable' Chronic Disease: An Analysis under Prospect Theory.* Soc Sci Med, 2018. **214**: p. 144-153.
28. Attema, A.E., J.J. Frasch, and O. L'Haridon, *Multivariate Risk Preferences in the Quality-Adjusted Life Year Model.* Health Econ, 2022. **31**(2): p. 382-398.
29. Seabury, S.A., et al., *Patients Value Metastatic Cancer Therapy More Highly Than Is Typically Shown through Traditional Estimates.* Health Aff (Millwood), 2012. **31**(4): p. 691-699.
30. Kahneman, D.J. and A. Tversky, *Prospect Theory: An Analysis of Decision under Risk.* Econometrica, 1979. **47**(2): p. 363-391.
31. Bleichrodt, H. and J. Miyamoto, *A Characterization of Quality-Adjusted Life-Years under Cumulative Prospect Theory.* Mathematics of Operations Research, 2003. **28**(1): p. 181-193.
32. Stiggelbout, A.M. and J.C. de Haes, *Patient Preference for Cancer Therapy: An Overview of Measurement Approaches.* J Clin Oncol, 2001. **19**(1): p. 220-30.
33. University of Southern California Dornsife: Center for Economic and Social Research, *Understanding America Study*. University of Southern California.
34. Li, Z., Rohde, Kirsten I. M, Wakker, Peter P, *Improving One's Choices by Putting Oneself in Others' Shoes – an Experimental Analysis.* Journal of Risk and Uncertainty, 2017. **54**(1): p. 1-13.
35. Holt, C.A. and S.K. Laury, *Risk Aversion and Incentive Effects.* American Economic Review, 2002. **92**(5): p. 1644-1655.
36. Drichoutis, A.C. and J.L. Lusk, *What Can Multiple Price Lists Really Tell Us About Risk Preferences?* Journal of Risk and Uncertainty, 2016. **53**(2): p. 89-106.
37. Chetan, D., et al., *Eliciting Risk Preferences: When Is Simple Better?* Journal of Risk and Uncertainty, 2010. **41**(3): p. 219-243.
38. Rosen, S., *The Value of Changes in Life Expectancy.* Journal of Risk and Uncertainty, 1988. **1**(3): p. 285-304.
39. Chai, T. and R.R. Draxler, *Root Mean Square Error (Rmse) or Mean Absolute Error (Mae)? – Arguments against Avoiding Rmse in the Literature.* Geosci. Model Dev., 2014. **7**(3): p. 1247-1250.
40. Phelps, C. and D. Lakdawalla, *Health Technology Assessment with Diminishing Returns to Health: The Generalized Risk-Adjusted Cost Effectectiveness (Grace) Approach.* Value in Health, forthcoming.
41. Hirth, R.A., et al., *Willingness to Pay for a Quality-Adjusted Life Year: In Search of a Standard.* Medical Decision Making, 2000. **20**(3): p. 332-342.
42. Neumann, P.J., J.T. Cohen, and M.C. Weinstein, *Updating Cost-Effectiveness — the Curious Resilience of the $50,000-Per-Qaly Threshold.* New England Journal of Medicine, 2014. **371**(9): p. 796-797.

43. Institute for Clinical Economic Review, *Oral Immunotherapy and Viaskin® Peanut for Peanut Allergy: Effectiveness and Value*. 2019, Institute for Clinical and Economic Review: Boston, MA.
44. Barsky, R.B., et al., *Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study.* The Quarterly Journal of Economics, 1997. **112**(2): p. 537-579.
45. Halek, M. and J.G. Eisenhauer, *Demography of Risk Aversion.* The Journal of Risk and Insurance, 2001. **68**(1): p. 1-24.
46. Noussair, C.N., S.T. Trautmann, and G. van de Kuilen, *Higher Order Risk Attitudes, Demographics, and Financial Decisions.* The Review of Economic Studies, 2014. **81**(1): p. 325-355.
47. Sanders, G.D., et al., *Recommendations for Conduct, Methodological Practices, and Reporting of Cost-Effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine.* Jama, 2016. **316**(10): p. 1093-103.
48. Van Wilder, L., et al., *A Comprehensive Catalogue of Eq-5d Scores in Chronic Disease: Results of a Systematic Review.* Quality of Life Research, 2019. **28**(12): p. 3153-3161.
49. Sulmasy, D.P., et al., *The Culture of Faith and Hope.* Cancer, 2010. **116**(15): p. 3702-3711.
50. Nurgat, Z.A., et al., *Patient Motivations Surrounding Participation in Phase I and Phase Ii Clinical Trials of Cancer Chemotherapy.* British Journal of Cancer, 2005. **92**(6): p. 1001-1005.
51. Bosk, C.L., *Obtaining Voluntary Consent for Research in Desperately Ill Patients.* Medical Care, 2002. **40**(9): p. V-64-V-68.
52. Food and Drug Administration. *Right to Try*. January 23, 2023; Available from: https://www.fda.gov/patients/learn-about-expanded-access-and-other-treatment-options/right-try.
53. Tarn, D.M., et al., *Primary Nonadherence to Statin Medications: Survey of Patient Perspectives.* Preventive Medicine Reports, 2021. **22**: p. 101357.
54. Langer, A.S., *Side Effects, Quality-of-Life Issues, and Trade-Offs: The Patient Perspective.* JNCI Monographs, 2001. **2001**(30): p. 125-129.
55. Smerecnik, C.M., et al., *A Systematic Review of the Impact of Genetic Counseling on Risk Perception Accuracy.* J Genet Couns, 2009. **18**(3): p. 217-28.
56. Centers for Disease Control and Prevention. *Genetic Counseling*. June 24, 2022; Available from: https://www.cdc.gov/genomics/gtesting/genetic_counseling.htm.
57. Einav, L., A. Finkelstein, and H. Williams, *Paying on the Margin for Medical Dare: Evidence from Breast Cancer Treatments.* American Economic Journal: Economic Policy, 2016. **8**(1): p. 52-79.
58. Cohen, A. and L. Einav, *Estimating Risk Preferences from Deductible Choice.* American Economic Review, 2007. **97**(3): p. 745-788.
59. Lakdawalla, D. and N. Sood, *Health Insurance as a Two-Part Pricing Contract.* Journal of Public Economics, 2013. **102**: p. 1-12.
60. Soares, M.O., *Is the Qaly Blind, Deaf and Dumb to Equity? Nice's Considerations over Equity.* British medical bulletin, 2012. **101**(1).
61. Franken, M., et al., *A Comparative Study of the Role of Disease Severity in Drug Reimbursement Decision Making in Four European Countries.* Health Policy, 2015. **119**(2): p. 195-202.
62. Johri, M. and O.F. Norheim, *Can Cost-Effectiveness Analysis Integrate Concerns for Equity? Systematic Review.* International Journal of Technology Assessment in Health Care, 2012. **28**(2): p. 125-132.
63. Ward, T., et al., *Incorporating Equity Concerns in Cost-Effectiveness Analyses: A Systematic Literature Review.* Pharmacoeconomics, 2022. **40**(1): p. 45-64.
64. Cookson, R., M. Drummond, and H. Weatherly, *Explicit Incorporation of Equity Considerations into Economic Evaluation of Public Health Interventions.* Health Economics, Policy and Law, 2009. **4**(2): p. 231-245.

65. Wailoo, A., A. Tsuchiya, and C. McCabe, *Weighting Must Wait.* Pharmacoeconomics, 2009. **27**(12): p. 983-989.

7. APPENDIX

7.1.Structural Assumptions

Consumers may have risk preferences over health-related quality of life (HRQoL). Therefore, we consider risky prospects $(q_1, p_1; \ldots; q_n, p_n)$ yielding health-related quality of life outcome $q_i \in Q$ with probability $p_i \in P = [0,1]$, where $\sum_{i=1}^{n} p_i = 1$. Degenerate prospects yield certain health-related quality of life outcomes. For simplicity, we will denote binary prospects of the form $(q_1, p_1; q_2, p_2)$, as $(q_1\ p_1\ q_2)$. The set Q is the interval [0,1] in $\mathbb{R}^+$containing 0 (a fixed unit of time in a health state equivalent to death) and 1 (a fixed unit of time in full health). The set of all prospects considered is [0, 1]. The relation $\succcurlyeq$ denotes "at least as preferred as" over prospects, with $\succ$, $\sim$ as strict preference and equivalence, respectively, derived as usual from $\succcurlyeq$. The preference relation satisfies the von Neumann-Morgenstern axioms (1947).[1] Hence, there exists a utility function *U*, assigning to each quality-of-life outcome *q* the utility *U*(*q*), so that the expectation of *U*, ($\sum_{i=1}^{n} p_i U(q_i)$ for the above lottery), governs the choices among lotteries over health-related quality of life outcomes. An expected utility function is unique up to a positive linear transformation $U' = \beta U + \tau$, where $\beta > 0$.

Definition 1: (50-50 Certainty Equivalents). Let $(q_1\ \frac{1}{2}\ q_2)$ denote an even-chance prospect between q_1 and q_2 health-related quality of life outcomes. A "50-50 certainty equivalent" is the outcome $q_{CE} \in Q$ such that the decision-maker is indifferent between q_{CE} and (q_1 ½ q_2), i.e., $q_{CE} \sim (q_1\ \frac{1}{2}\ q_2)$.

It is worth noting that expected utility assumes invariance among certainty equivalents and other equivalent matching exercises (e.g., probability, gain and loss equivalents for the choice $q_{CE} \sim (q_1\ \frac{1}{2}\ q_2)$.

Definition 2: (Power-Risk Constancy). A decision-maker will be called power-risk constant if, for $q_{CE}, q_1, q_2 > 0$ and probability $p = \frac{1}{2}$, we have the indifference relationship:

$$(q_{CE}^{a} + z^{a})^{-a} \sim \left((q^{a_1} + z^{a})^{-a}\ \frac{1}{2}\ (q^{a_2} + z^{a})^{-a}\right), \qquad \text{(i)}$$

where $a \neq 0$, and the condition $q_{CE}^{a}, q^{a_1}, q^{a_2}, (q_{CE}^{a} + z^{a})^{-a}, (q_1 + z^{a})^{-a}, (q_1 + z^{a})^{-a}, (q_1 + z^{a})^{-a}, (q_2 + z^{a})^{-a} \in Q = [0,1]$ is satisfied uniformly in *z*.

Theorem 1. *The expected utility function of a power-risk constant decision-maker (Def. 2) is an expo-power function (Eq. 1a).*

Proof: Aczél showed that for any continuous group arithmetic operation $\circ$,where $q \circ q'$ is defined on the open interval $Q = (0,1)$, there exists a scaling function $f(\cdot)$ such that $q \circ q' = f^{-1}(f(q) + f(q'))$, with $f(\cdot)$ strictly increasing and continuous on $Q = (0,1)$.[2, p. 254] The scaling function translates $\circ$ into $+$. Aczél [2, Theorem 3.1.3.2], Harvey [3, Theorem 1b], Miyamoto [4, Lemma 1], Nagumo [5, p. 78 and proof] and Pfanzagl [6, Theorem 3, p.290] guarantee that: $q_{CE} \sim (\mathrm{q}_1 \, \frac{1}{2} \, q_2)$ if and only if $q_{CE} + z \sim (\mathrm{q}_1 + z \, \frac{1}{2} \, q_2 + z)$, implies an exponential utility function, $U(q) = be^{rq} + c$ where $b > 0$ and $r \neq 0$ and $U(q) = bq + c$ where $b > 0$ and $r = 0$. From Aczél [2, p. 254], we can define a power-risk constancy operation $q \circ \mathrm{z} = (q^a + z^a)^{-a}$. Consider the scaling function of power form $f(x) = x^a$, where $a \neq 0$. Since the scaling function $f(\cdot)$ translates $\circ$ into $+$, power-risk constancy implies $U(q) = be^{rf(q)} + c$ where $b > 0$ and $r \neq 0$, and $U(q) = bf(q) + c$ where $b > 0$ and $r = 0$ (see [3, p. 1482, Eq. 3]). Since $f(q)$ is a power scaling function, with power not equal to 0, the utility function is expo-power as given in Eq. 1a. It is left to show that the limits of U as q approaches 0 and 1, respectively, are real. $\lim_{q \to 0} U(q) = c + 1$, and $\lim_{q \to 1} U(q) = be^r + c$, which reduces to $b + c$ for the case where $r = 0$. Having shown the existence of real limits at 0 and 1, we set $U(0)$ and $U(1)$ equal to their real limits. □

7.2. Additional results: pooled utility and risk estimates

Table 1. Risk preference parameter equations

	Relative risk aversion $r_H = -\frac{W''(H)}{W'(H)}H$	Relative risk prudence $\pi_H = -\frac{W'''(H)}{W''(H)}H$
CRRA	ρ	$\rho + 1$
Expo-power	$abH^a - a + 1$	$(2-a) + abH^a - \frac{a^2 bH^a}{1 - a + abH^a}$

CRRA utility only has one parameter (ρ), and its estimated variance equals 0.00166006. The variance-covariance matrix for expo-power is provided in Table 8.

Table 2. Pooled expo-power utility estimates, variance covariance matrix

	a	b
a	0.00223376	
b	0.00378934	0.00835446

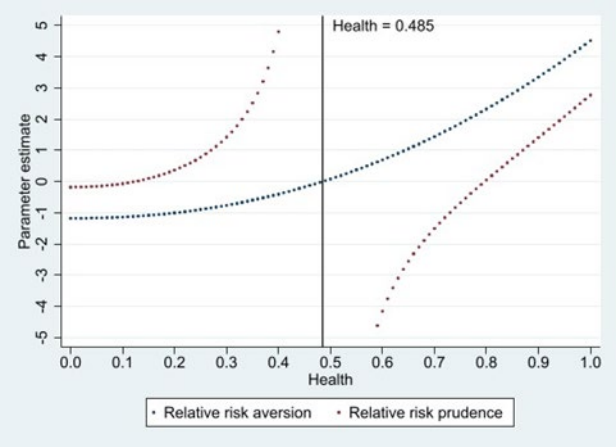
Notes: Variance covariance matrix of estimated expo-power utility estimates (Table 3) from pooled estimation.

Table 3. Risk preference parameter estimates at selected values of health, pooled estimates

	Expo-power		CRRA	
	r_H^*	π_H^*	r_H^*	π_H^*
H = 0.1	-1.14 [-1.267, -1.005]	-0.07 [-0.194, 0.062]	0.2822 [0.217, 0.349]	1.2822 [1.217, 1.349]
H = 0.5	-0.08 [-0.037, 0.198]	-31.83 [-179.1, 98.3]		
H = 0.9	3.35 [2.846, 3.917]	1.41 [1.027, 1.835]		
H = 1.0	4.51 [3.820, 5.332]	2.77 [2.239, 3.403]		
Estimation N	9,710		9,710	

Notes: Estimation sample pooled all 13 health gamble questions for N=747 respondents. 95% confidence intervals for the risk preference parameters generated using 1,000 bootstrap replications clustered at the respondent level. Standards errors are similar without clustering.

Figure 1. Expo-power utility relative risk parameters, pooled estimates



Notes: Relative risk aversion and prudence calculated from utility parameter estimates for the expo-power utility function (Table 2). Preferences switch from risk loving to risk averse at a health value of 0.485. This value corresponds to an asymptote for prudence, which does not exist when risk aversion equals zero.

7.3. Additional Results: Individual risk estimates and regressions

Table 4. Respondent summary statistics (self-rated health subgroups)

	All respondents [N=747]	Self-rated health = 86-100 [N=218]	Self-rated health = 76-85 [N=262]	Self-rated health = 0-75 [N=267]
Male	0.420 (0.493)	0.380 (0.486)	0.458 (0.499)	0.415 (0.493)
Age (years)	52.15 (16.06)	49.29 (16.79)	53.42 (15.78)	53.24 (15.48)
Age group: <40 years	0.259 (0.438)	0.339 (0.474)	0.221 (0.415)	0.232 (0.423)
Age group: 40-54	0.270 (0.444)	0.247 (0.432)	0.297 (0.458)	0.262 (0.440)
Age group: 55-64	0.196 (0.397)	0.183 (0.387)	0.183 (0.387)	0.220 (0.415)
Age group: 65+	0.273 (0.445)	0.229 (0.421)	0.297 (0.458)	0.284 (0.452)
Married	0.574 (0.494)	0.605 (0.489)	0.595 (0.491)	0.528 (0.500)
College graduate	0.527 (0.499)	0.646 (0.479)	0.538 (0.499)	0.419 (0.494)
White, non-Hispanic	0.709 (0.454)	0.711 (0.454)	0.721 (0.449)	0.696 (0.460)
Working	0.566 (0.495)	0.688 (0.464)	0.580 (0.494)	0.453 (0.498)
Income: <60K	0.389 (0.487)	0.288 (0.454)	0.339 (0.474)	0.520 (0.500)
Income: 60K-99.9K	0.293 (0.455)	0.288 (0.454)	0.343 (0.475)	0.247 (0.432)

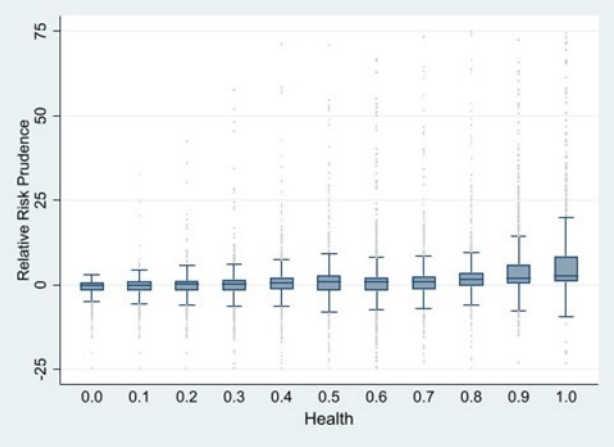
Income: 100K+	0.317 (0.465)	0.422 (0.495)	0.316 (0.466)	0.232 (0.423)
Insured	0.716 (0.451)	0.688 (0.464)	0.755 (0.430)	0.700 (0.458)
Region: Northwest	0.133 (0.340)	0.165 (0.372)	0.083 (0.277)	0.157 (0.364)
Region: Midwest	0.231 (0.422)	0.206 (0.405)	0.293 (0.456)	0.191 (0.393)
Region: South	0.269 (0.443)	0.302 (0.460)	0.232 (0.423)	0.277 (0.448)
Region: West	0.364 (0.481)	0.325 (0.469)	0.389 (0.488)	0.370 (0.483)
Self-rated health: 0-75	0.357 (0.479)	0.0 (0.0)	0.0 (0.0)	1.0 (0.0)
Self-rated health: 76-85	0.350 (0.477)	0.0 (0.0)	1.0 (0.0)	0.0 (0.0)
Self-rated health: 86-100	0.291 (0.454)	1.0 (0.0)	0.0 (0.0)	0.0 (0.0)

Table 5. Individual expo-power utility and risk parameter estimates, summary statistics

	Median [IQR]		Mean (SD)	
Utility parameters				
a	2.08 [1.27, 3.65]		3.44 (5.10)	
b	2.99 [0.89, 6.77]		10,555 (148,317)	
Risk preference parameters evaluated at select values for health (H)				
	r_H^*	π_H^*	r_H^*	π_H^*
H = 0.1	-1.00 [-2.65, -0.10]	-0.14 [-1.80, 0.93]	-2.32 (5.17)	-1.39 (5.98)
H = 0.5	0.25 [-1.03, 1.13]	0.90 [-1.90, 2.59]	1.39 (19.7)	2.21 (24.5)
H = 0.9	3.26 [0.82, 11.0]	2.31 [0.78, 8.75]	8,719 (109,357)	8,715 (109,355)
H = 1.0	4.36 [0.93, 18.7]	3.49 [1.24, 16.3]	383,324 (5,540,996)	383,322 (5,540,996)
N	747			

Notes: Results generated from estimation of expo-power utility models for each individual using all 13 health gamble questions.

Figure 2. Individual relative risk prudence parameter estimates for full range of health [0,1]



Notes: Distribution of risk preference parameters over the full range of health. The IQR is represented by the upper and lower edges of the rectangles. The whiskers represent the upper and lower adjacent values, which are calculated as the 75th percentile + 1.5*IQR and 25th percentile + 1.5*IQR, respectively. Individual prudence parameters were derived from expo-power utility for the majority of respondents (N=639). The remaining N=108 respondents had expo-power utility parameter estimates that imply CRRA preferences and a smaller RMSE value for the CRRA utility model.

Table 6. Individual risk preference classifications

	Risk seeking over full range of health	Switcher	Risk averse over full range of health
DARA-DRRA	N.F.	0	4
DARA-CRRA	78	N.F.	7
DARA-IRRA	N.F.	N.F.	56
CARA-IRRA	N.F.	N.F.	0
IARA-CRRA	1	N.F.	59
IARA-IRRA	15	527	N.F.

Note: N.F.=not feasible. All possible absolute-relative risk structure combinations are included in the table. Switchers can either switch from risk seeking to risk averse (feasible for IARA-IRRA) or risk averse to risk seeking (feasible for DARA-DRRA).

Figure 3. Histogram of switch points (risk seeking to risk averse)

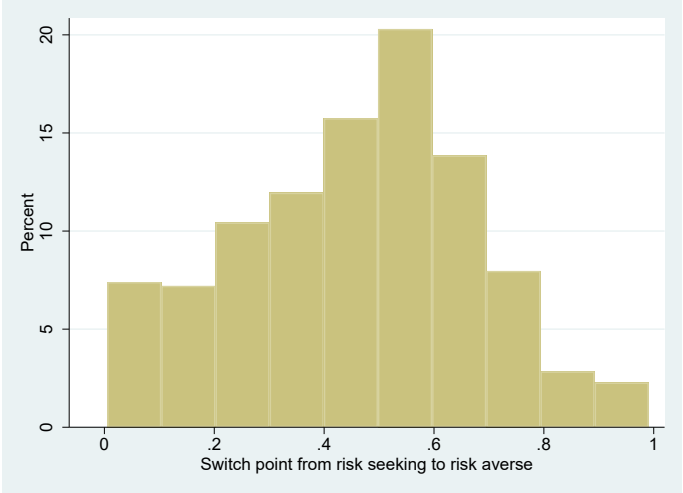


Table 7. Effect of demographics on risk type, regression results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Dep. Variable = 1 if CRRA, 0 if IRRA					**Dep. Variable = 1 if DARA, 0 if IARA**				
Self-rated health: 76-85	0.0277 (0.0376)	0.174 (0.242)				0.0470 (0.0356)	0.342 (0.266)			
Self-rated health: 86-100	-0.000799 (0.0378)	-0.0105 (0.250)				0.0396 (0.0361)	0.283 (0.271)			
Income: $60K-99.9K	0.0338 (0.0393)	0.224 (0.257)	1.079** (0.509)	-0.118 (0.436)	-0.213 (0.461)	-0.0875** (0.0367)	-0.611** (0.264)	0.394 (0.496)	-0.687 (0.453)	-1.379*** (0.517)
Income: $100K+	0.0490 (0.0444)	0.327 (0.287)	1.504*** (0.564)	-0.130 (0.466)	-0.163 (0.444)	-0.0602 (0.0391)	-0.378 (0.264)	0.735 (0.545)	-0.524 (0.417)	-1.150*** (0.440)
Insured	-0.0530 (0.0347)	-0.333 (0.207)	-0.595 (0.433)	-0.249 (0.353)	-0.450 (0.378)	-0.0238 (0.0311)	-0.163 (0.217)	-0.360 (0.444)	-0.373 (0.365)	0.0406 (0.399)
Age group: 40-54	0.0118 (0.0409)	0.0741 (0.262)	-0.0307 (0.470)	0.130 (0.443)	0.166 (0.478)	-0.00454 (0.0368)	-0.0274 (0.284)	-0.170 (0.531)	-0.362 (0.487)	0.480 (0.573)
Age group: 55-64	-0.0283 (0.0450)	-0.196 (0.307)	-0.724 (0.636)	-0.101 (0.537)	-0.129 (0.555)	0.0224 (0.0417)	0.172 (0.298)	-0.883 (0.668)	0.723 (0.531)	0.336 (0.566)
Age group: 65+	0.0101 (0.0527)	0.0622 (0.333)	-0.383 (0.638)	0.314 (0.581)	0.210 (0.586)	0.0245 (0.0444)	0.158 (0.329)	-0.310 (0.561)	0.479 (0.618)	0.380 (0.623)
Male	0.0275 (0.0313)	0.176 (0.198)	0.537 (0.463)	0.110 (0.329)	0.0406 (0.348)	0.0198 (0.0292)	0.142 (0.207)	0.199 (0.469)	0.0526 (0.348)	0.226 (0.368)
Married	-0.0139 (0.0321)	-0.0868 (0.203)	0.447 (0.387)	-0.875*** (0.333)	0.411 (0.355)	-0.0100 (0.0289)	-0.0701 (0.208)	0.439 (0.412)	-0.467 (0.342)	0.108 (0.380)
College graduate	0.0262 (0.0338)	0.169 (0.217)	0.434 (0.455)	-0.334 (0.346)	0.422 (0.353)	0.0527* (0.0295)	0.398* (0.226)	0.438 (0.490)	-0.208 (0.338)	0.934** (0.395)
Employed	0.00306 (0.0385)	0.0250 (0.248)	0.375 (0.482)	0.109 (0.434)	-0.443 (0.458)	0.00122 (0.0320)	0.0113 (0.240)	0.00296 (0.465)	0.518 (0.427)	-0.644 (0.420)
White, non-Hispanic	0.0445 (0.0348)	0.296 (0.234)	0.806 (0.490)	0.466 (0.382)	-0.121 (0.416)	0.0312 (0.0327)	0.244 (0.248)	0.419 (0.534)	0.131 (0.410)	0.132 (0.461)
Region: Midwest	-0.0837 (0.0538)	-0.503 (0.314)	-0.586 (0.672)	-0.488 (0.562)	-0.237 (0.516)	-0.112** (0.0496)	-0.825** (0.347)	-0.717 (0.663)	-1.032 (0.630)	-0.593 (0.614)
Region: South	-0.0837 (0.0520)	-0.507* (0.300)	0.142 (0.559)	-0.287 (0.577)	-1.059** (0.514)	-0.0695 (0.0497)	-0.470 (0.317)	-0.581 (0.609)	-0.0575 (0.597)	-0.576 (0.546)
Region: West	-0.0450 (0.0514)	-0.246 (0.283)	0.446 (0.568)	-0.136 (0.567)	-0.811* (0.455)	-0.0328 (0.0497)	-0.206 (0.295)	0.0299 (0.583)	0.0141 (0.560)	-0.650 (0.511)
Constant	0.202** (0.0790)	-1.454*** (0.486)	-3.421*** (0.998)	-0.733 (0.918)	-0.660 (0.761)	0.198*** (0.0720)	-1.533*** (0.504)	-2.231** (1.021)	-0.871 (0.808)	-1.044 (0.861)
Sample	All respondents		Self-rated health: 86-100	Self-rated health: 76-85	Self-rated health: 0-75	All respondents		Self-rated health: 86-100	Self-rated health: 76-85	Self-rated health: 0-75
N	742	742	216	262	264	746	746	218	262	266
Estimation	OLS	Logit	Logit	Logit	Logit	OLS	Logit	Logit	Logit	Logit

Notes: All regressions exclude 1 individual who did not have complete data for demographic variables. Relative risk regressions exclude N=4 people with DRRA preferences. No individuals had CARA preferences. Clustered standard errors shown in parenthesis. Results are similar if we estimate probit models. Reference categories: self-rated health 0-75, income <$60K, age less than 40, Northeast region. Significance: ***p<0.01; **p<0.05; *p<0.1.

Table 8. Regression results: Effect of demographics on relative risk aversion at select values of health, subsample with expo-power utility

	Self-rated health = 86-100 [N=190]			Self-rated health = 76-85 [N=220]			Self-rated health = 0-75 [N=228]		
	H=0.2	H=0.5	H=1.0	H=0.2	H=0.5	H=1.0	H=0.2	H=0.5	H=1.0
Income: $60K-99.9K	0.492	0.157	0.064	-0.407	-0.320	-0.954	-0.909**	-0.572	3.201
	(0.388)	(0.426)	(2.707)	(0.413)	(0.418)	(2.376)	(0.440)	(0.445)	(4.153)
Income: $100K+	0.743	1.022**	1.050	-0.723	-0.404	-0.980	-0.749*	-0.425	1.600
	(0.456)	(0.501)	(4.329)	(0.503)	(0.491)	(2.745)	(0.389)	(0.466)	(3.267)
Insured	0.153	0.493	1.360	-0.871**	-0.328	0.829	-0.519	-0.304	0.342
	(0.476)	(0.490)	(2.834)	(0.408)	(0.407)	(2.534)	(0.417)	(0.400)	(3.821)
Age group: 40-54	-0.467	-0.379	2.716	-1.764***	-0.411	3.051	-0.360	0.124	7.280
	(0.529)	(0.538)	(2.847)	(0.591)	(0.499)	(3.548)	(0.627)	(0.496)	(5.109)
Age group: 55-64	-0.525	-0.176	2.160	-0.502	-0.488	0.447	-0.013	0.693	4.117
	(0.581)	(0.584)	(3.778)	(0.517)	(0.592)	(2.922)	(0.478)	(0.511)	(3.423)
Age group: 65+	-0.289	-0.275	-1.432	-0.297	0.181	0.587	0.087	1.138**	4.392
	(0.663)	(0.669)	(5.045)	(0.668)	(0.662)	(3.839)	(0.465)	(0.507)	(4.369)
Male	0.228	-0.305	-2.262	0.095	0.031	-2.259	0.492	-0.242	-3.237
	(0.457)	(0.383)	(2.142)	(0.346)	(0.311)	(1.876)	(0.362)	(0.335)	(2.619)
Married	0.405	0.438	-0.208	0.504	-0.041	0.288	-0.161	0.100	-0.186
	(0.394)	(0.380)	(1.916)	(0.371)	(0.366)	(1.860)	(0.358)	(0.380)	(2.659)
College graduate	0.652	0.315	-3.919	0.863**	0.017	-6.850**	0.893**	0.218	-6.010*
	(0.520)	(0.489)	(3.934)	(0.398)	(0.366)	(2.894)	(0.348)	(0.392)	(3.387)
Currently working	0.544	-0.222	-4.194	0.061	-0.217	-0.744	-0.237	0.039	1.163
	(0.508)	(0.501)	(4.455)	(0.539)	(0.556)	(3.077)	(0.354)	(0.450)	(3.240)
White, non-Hispanic	0.708	0.571	-1.958	0.337	0.454	-0.216	0.726*	-0.660	-6.291
	(0.459)	(0.521)	(2.962)	(0.525)	(0.413)	(2.023)	(0.396)	(0.472)	(4.858)
Region: Midwest	-0.036	-0.118	-0.933	-0.071	-1.569*	1.541	0.331	0.618	-2.336
	(0.527)	(0.502)	(3.054)	(0.712)	(0.923)	(7.221)	(0.764)	(0.613)	(7.041)
Region: South	-0.513	0.066	5.581	0.508	-1.260	-1.480	0.235	0.416	0.697
	(0.566)	(0.581)	(4.612)	(0.767)	(0.913)	(7.257)	(0.692)	(0.545)	(7.269)
Region: West	-0.158	-0.061	-0.574	0.320	-1.008	0.746	0.281	0.456	-0.959
	(0.491)	(0.521)	(3.021)	(0.740)	(0.922)	(7.048)	(0.633)	(0.549)	(6.601)
Constant	-2.568***	-0.827	11.700	-0.885	1.671	11.179	-1.208	0.615	12.831
	(0.841)	(0.858)	(7.398)	(1.069)	(1.133)	(9.192)	(0.870)	(0.878)	(9.856)

Notes: Results are estimated coefficients from a median regression with the dependent variable equal to r_H^* evaluated at the health level indicated in the 2nd row (0.2, 0.5, or 1.0). Estimation sample includes the N=638 respondents with expo-power utility. Clustered standard errors shown in parenthesis. Reference categories: self-rated health 0-75, income <$60K, age less than 40, Northeast region. Significance: ***p<0.01; **p<0.05; *p<0.1

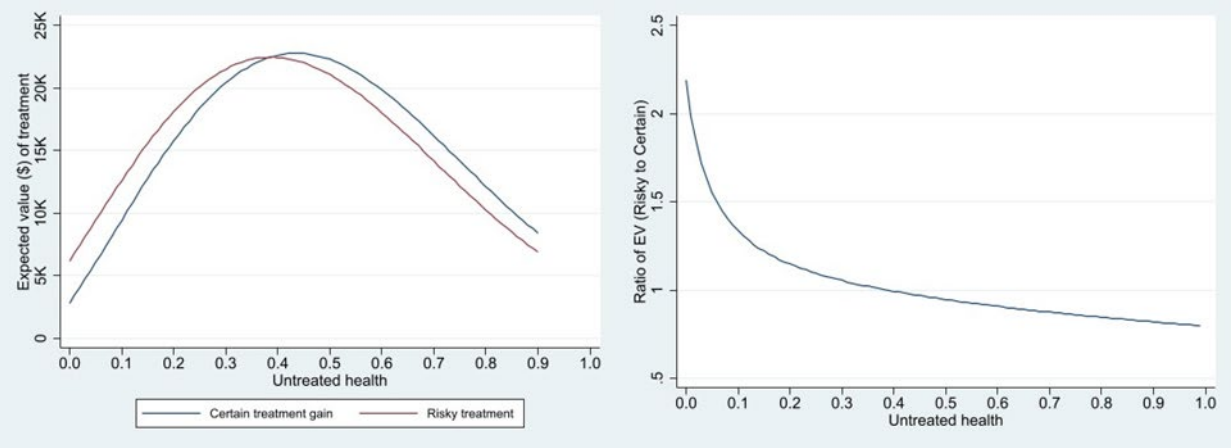
Table 9. Regression results: Effect of demographics on relative risk aversion, subsample with CRRA utility

	Self-rated health = 86-100 [N=28]	Self-rated health = 76-85 [N=42]	Self-rated health = 0-75 [N=38]
Income: $60K-99.9K	1.572 (4.121)	-1.325 (1.577)	-1.514 (2.305)
Income: $100K+	2.023 (3.551)	0.006 (1.639)	-2.513 (2.681)
Insured	0.870 (3.895)	-0.750 (1.337)	1.558 (2.194)
Age group: 40-54	-1.090 (5.321)	-0.879 (1.591)	-0.989 (2.506)
Age group: 55-64	-3.428 (4.602)	0.108 (1.642)	0.082 (2.969)
Age group: 65+	-1.865 (7.915)	-0.004 (2.162)	-2.177 (2.863)
Male	-1.546 (4.404)	0.125 (1.085)	0.955 (1.844)
Married	2.861 (3.176)	0.761 (1.413)	-0.805 (2.100)
College graduate	0.065 (2.954)	0.323 (1.114)	0.868 (2.433)
Currently working	-2.883 (8.165)	1.238 (1.734)	-2.537 (1.806)
White, non-Hispanic	-4.236 (5.995)	-0.029 (1.490)	0.113 (1.652)
Region: Midwest	-1.024 (7.831)	0.191 (1.825)	1.077 (3.132)
Region: South	-4.196 (4.839)	0.511 (1.871)	1.274 (2.814)
Region: West	-5.089 (5.960)	-0.155 (1.707)	1.448 (2.574)
Constant	5.511 (13.200)	-0.698 (2.885)	-0.177 (4.565)

Notes: Results are estimated coefficients from a median regression with the dependent variable equal to r_H^*. Estimation sample includes the N=108 respondents with CRRA utility. If we estimate a regression with all self-rated health groups pooled (to increase sample size), all coefficients remain statistically insignificant. Clustered standard errors shown in parenthesis. Reference categories: self-rated health 0-75, income <$60K, age less than 40, Northeast region. Significance: ***p<0.01; **p<0.05; *p<0.1

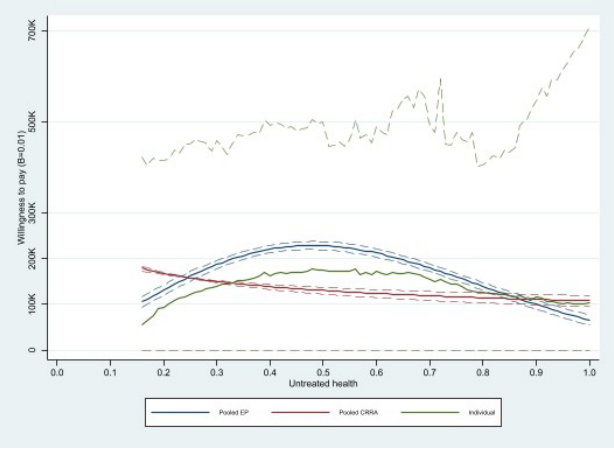
7.4. Additional results: implications for cost-effectiveness analysis

Figure 4. Expected value of two hypothetical treatments



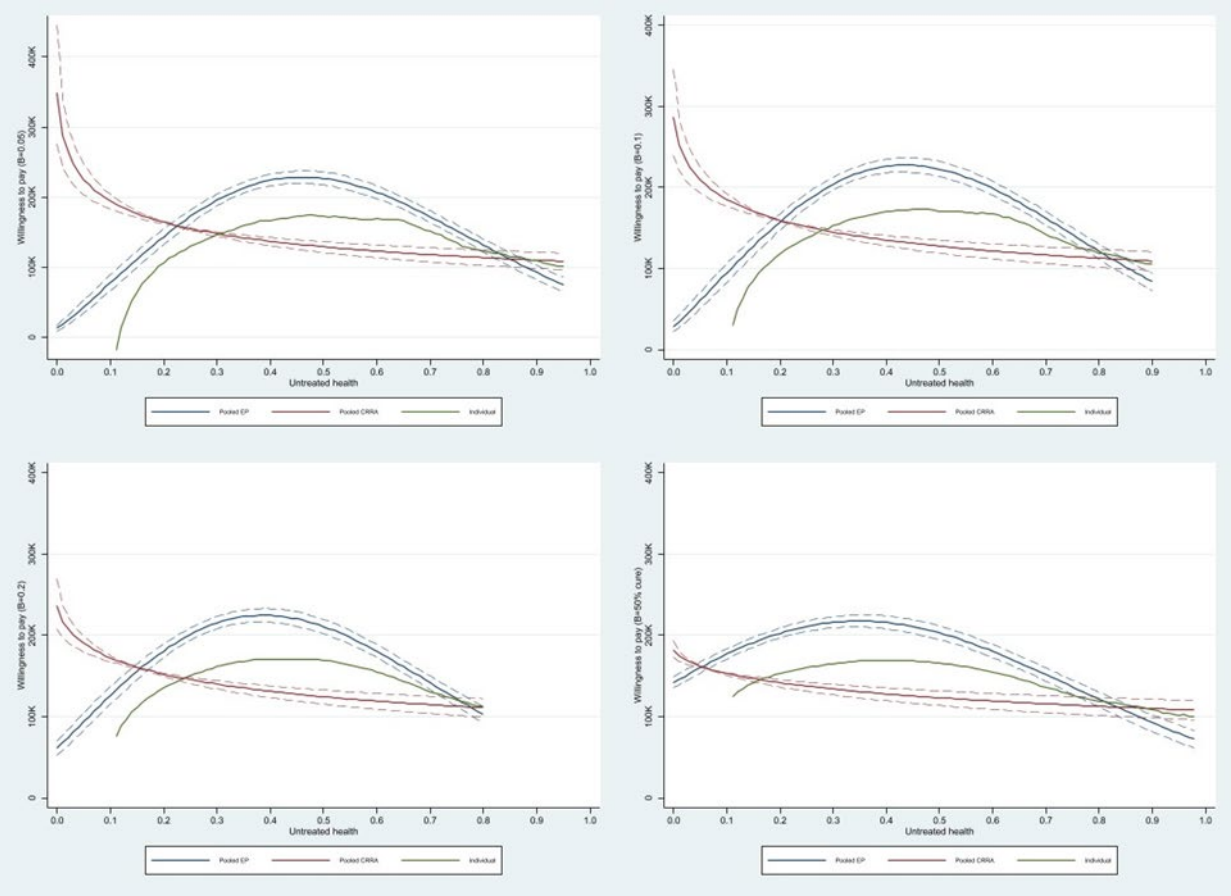
Notes: Expected value calculated using an expo-power utility function. The certain treatment has B = 0.1 with probability 1. The risky treatment has a 50% chance of B=0.2 and a 50% B=0 (i.e., no HRQoL benefit).

Figure 5. Comparison of pooled and individual WTP (with individual 95% confidence intervals)



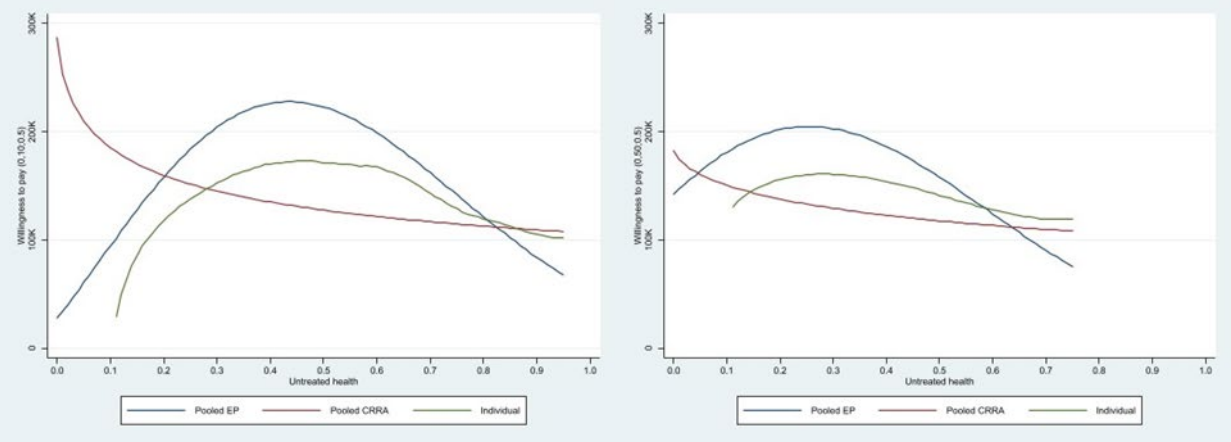
Notes: This figure reproduces Figure 5 and adds 95% CIs around the aggregate individual WTP curve. Individual curve corresponds to the per capita risk adjusted WTP derived from the individual level utility estimates summarized in Table 5. Dashed lines correspond to the 95% confidence intervals (CI).

Figure 6. Comparison of pooled and individual WTP corresponding to different treatment benefits



Notes: All WTP values assume an unadjusted K equal to $150,000. Individual curve corresponds to the per capita risk adjusted WTP derived from the individual level utility estimates summarized in Table 5. Dashed lines correspond to the 95% confidence intervals (CI); CIs for the individual WTP are wider than the y-axis scale and therefore omitted. Treatment benefit sizes equal 0.05 QoL burden reduction (top left panel); 0.1 QoL burden reduction (top right panel); 0.2 QoL burden reduction (bottom left panel); and a 50% QoL burden reduction (bottom right panel).

Figure 7. Comparison of pooled and individual WTP corresponding to risky treatments



Notes: All WTP values assume an unadjusted K equal to $150,000. Individual curve corresponds to the per capita risk adjusted WTP derived from the individual level utility estimates summarized in Table 5. Left panel: risky treatment has a 50% chance of

no benefit and 50% chance of 0.1 QoL benefit reduction. Right panel: risky treatment has a 50% chance of no benefit and 50% chance of 0.5 QoL benefit reduction.

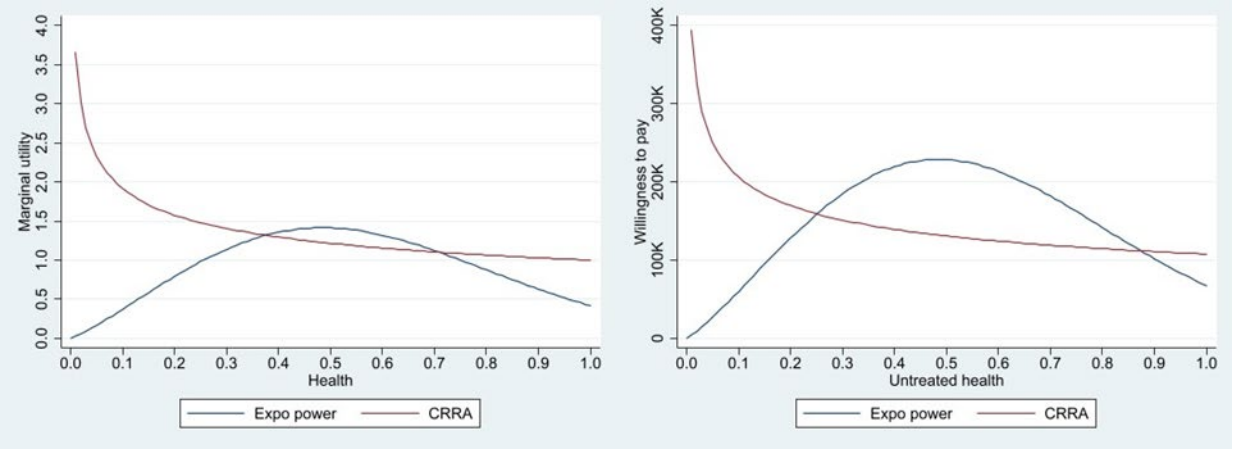
7.5. Comparison of K_{GRACE} across utility functions

We now turn to the relevance of functional forms for utility. K_{GRACE} depends on the level of utility at perfect health and the curvature of the utility function evaluated at health in the sick state. Individuals are risk-averse at perfect health under both expo-power and CRRA utility. However, the two functions differ in both their qualitative and quantitative predictions for risk preferences. To demonstrate the impact of the difference in underlying utility functions on GRACE, Table 16 provides a comparison of K_{GRACE} thresholds for expo-power and CRRA utility at selected health values, and Figure 10 shows the relationship between marginal utility and K_{GRACE}. While CRRA and expo-power utility yield similar K_{GRACE} for an illness associated with relatively high levels of health (0.9), their corresponding K_{GRACE} estimates diverge quickly thereafter. Under expo-power, traditional CEA would underestimate WTP for treatments for illnesses that cause health to fall between 0.24 and 0.77. The risk-seeking preferences at low levels of health for expo-power imply that traditional CEA will overestimate WTP at levels of health that are quite severe (0 to 0.23). If we introduce risk in treatment outcomes (Figure 6, left panel) CEA can underestimate WTP for treatments for any illness with health between 0 and 0.55. In contrast, under CRRA utility, traditional CEA would underestimate WTP for treatments for the most severe illnesses (i.e., health between 0 and 0.3).

Table 10. Risk-adjusted WTP thresholds under different utility functions (select health values)

Untreated health	Unadjusted K = $50K		Unadjusted K = $150K	
	Expo-power	CRRA	Expo-power	CRRA
0.90	33,904	36,973	101,711	110,919
0.75	54,059	38,925	162,178	116,776
0.50	76,170	43,644	228,509	130,932
0.25	52,904	53,073	158,711	159,219
0.10	20,117	68,734	60,350	206,203

Figure 8. Marginal utility and risk-adjusted WTP thresholds under different utility functions



Notes: Marginal utilities and risk-adjusted WTP calculated using the estimated utility parameters for expo power and CRRA (Table 3). Risk-adjusted WTP calculated assuming unadjusted K = $150,000.

7.6.HARA Utility

The GRACE theoretical framework was presented using the hyperbolic absolute risk-aversion (HARA) utility function, which nests exponential, power, linear, quadratic, and logarithmic utility [7]. Like expo-power utility, HARA nests constant, increasing, and decreasing relative risk-aversion. However, unlike expo power, HARA does not permit risk-seeing behavior. HARA utility takes the form:

$$W^{HARA}(q) = \frac{1-\gamma}{\gamma}\left(\frac{aq}{1-\gamma} + b\right)^{\gamma} \quad \text{(ii)}$$

To ensure monotonicity, we require $\left(\frac{aq}{1-\gamma} + b\right) > 0$ and $0 < \gamma < 1$. The first of these two constraints also implies concavity, and therefore rules out risk-seeking behavior. Following a similar approach to that outlined for expo power utility, we can write the HARA estimating equation as:

$$q_{CE}^{HARA} = \frac{1-\gamma}{a}\left[-b + \sqrt[\gamma]{0.5\left(\frac{aq_1}{1-\gamma}\right)^{\gamma} + 0.5\left(\frac{aq_2}{1-\gamma}\right)^{\gamma}}\right] \quad \text{(iii)}$$

Because a, b, and γ cannot be separately identified, we normalize a=1.[1] [8, 9]

We could not obtain pooled estimates for the HARA function due to non-convergence. We hypothesize this occurred because HARA rules out risk-seeking behavior and our sample appears to

[1] HARA utility requires b=0 to satisfy the Rosen criteria (i.e., U(0)=0). However, this restriction results in HARA collapsing to the CRRA case. To allow other risk preference structures and satisfy the Rosen criteria, one could use the estimated values for b and γ to create a "shifted HARA" function, defined as $W^{HARA}(q) + \frac{1-\gamma}{\gamma}b^{\gamma}$.

display some risk-seeking preferences. To test this hypothesis, we first estimated HARA utility on the subsample of individuals with estimated utility parameters consistent with risk-aversion over the full range of health. We then incrementally added subsets of people with small degrees of risk seeking behavior (i.e., have a switch point from risk seeking to risk averse for health less than 0.05, less than 0.1, and so on) to identify the point at which HARA no longer converges.

HARA estimates converged for the subsample of individuals (N=118) who were risk-averse over the full range of health.[2] Of note, the estimated value of b was approximately 0 (<0.000001), which implies HARA simplifies to the CRRA case. Furthermore, the estimated value of γ equaled 0.0571, which would yield implausible WTP values under GRACE [10]. The addition of the subset of people with risk seeking behavior for levels of health < 0.05 resulted in non-convergence. As a result, we conclude HARA utility should not be used in our full sample.

1. Von Neumann, J. and O. Morgenstern, *Theory of Games and Economic Behavior*. Second Edition ed. 1947, Princeton, New Jersey: Princeton University Press.
2. Aczél, J., *Lectures on Functional Equations and Their Applications*. 1966, New York: Academic Press.
3. Harvey, C.M., *Structured Prescriptive Models of Risk Attitudes.* Management Science, 1990. **36**(12): p. 1479-1501.
4. Miyamoto, J.M., *An Axiomatization of the Ratio/Difference Representation.* Journal of Mathematical Psychology, 1983. **27**(4): p. 439-455.
5. Nagumo, M., *Über Eine Klasse Der Mittelwerte.* Journal of Mathematics, 1930. **7**: p. 71-79.
6. Pfanzagl, J., *A General Theory of Measurement Applications to Utility.* Naval research logistics quarterly, 1959. **6**(4): p. 283-294.
7. Lakdawalla, D.N. and C.E. Phelps, *Health Technology Assessment with Risk Aversion in Health.* Journal of Health Economics, 2020. **72**.
8. Meyer, J., *Representing Risk Preferences in Expected Utility Based Decision Models.* Annals of Operations Research, 2010. **176**: p. 179-190.
9. Brocas, I., et al., *Risk Aversion in a Dynamic Asset Allocation Experiment.* Journal of financial and quantitative analysis, 2019. **54**(5): p. 2209-2232.
10. Phelps, C.E. and C. Cinatl, *Estimating Optimal Willingness to Pay Thresholds for Cost-Effectiveness Analysis: A Generalized Method.* Health Economics, 2021. https://doi.org/10.1002/hec.4268.

[2] Because the estimating equation for HARA is highly nonlinear, the nonlinear least square command in Stata (nl) did not converge unless we relaxed the convergence criteria substantially. We therefore utilized the menl command, which nests nonlinear least squares in a maximum likelihood environment and therefore has improved capacity for handling nonlinearities. Of note, the HARA function did not converge using menl for the full sample.