The Development of a Preference-based Scoring System for PROMIS® (PROPr): A Technical Report

v1.4

December 2017

Prepared by Janel Hanmer and Barry Dewitt

Preface

This technical report has been prepared as a reference document to record the details of the development of the PROMIS-Preference (PROPr) score. PROPr is a preference-based scoring system for seven PROMIS domains created using multiplicative multi-attribute utility theory. It serves as a generic, societal, preference-based summary scoring system of health-related quality of life.

This document contains a comprehensive overview of scoring system design, estimation, and usage recommendations. Standardized code for calculating PROPr scores using SAS and R are provided in the appendixes.

Acknowledgements

Collaborators

This work was undertaken with many generous and thoughtful collaborators:

- David Cella, PhD, Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, Il. D-cella@northwestern.edu
- David Feeny, PhD, Department of Economics, McMaster University, Hamilton, ON, Canada; Health Utilities Incorporated, Dundas, ON, Canada.
- Baruch Fischhoff, PhD, Department of Engineering and Public Policy and Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA.
- William J. Furlong, McMaster University, Hamilton, ON, Canada; Health Utilities Incorporated, Dundas, ON, Canada.
- Ron D. Hays, PhD, Division of General Internal Medicine & Health Services Research, UCLA, Los Angeles, CA.
- Rachel Hess, MD, MS, Division of Health System Innovation and Research, University of Utah Schools of the Health Sciences, Salt Lake City, UT.
- Paul A Pilkonis, PhD, Department of Psychiatry, University of Pittsburgh Medical Center, Pittsburgh, PA.
- Dennis Revicki, PhD, Outcomes Research, Evidera, Bethesda, MD.
- Mark Roberts MD, MPP, Department of Health Policy and Management – University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA.
- Joel Tsevat, MD, MPH, Division of General Internal Medicine, University of Cincinnati College of Medicine and Cincinnati VA Medical Center, Cincinnati, OH.
- Lan Yu, PhD, Department of General Internal Medicine, University of Pittsburgh Medical Center, Pittsburgh, PA.

Funding

Janel Hanmer was supported by the National Institutes of Health through Grant Number KL2 TR001856. Barry Dewitt is partially supported by a Social Sciences and Humanities Research Council of Canada Doctoral Fellowship.

Participant recruitment for domain selection and valuation method testing was completed by the Clinical and Translational Science Institute at the University of Pittsburgh, which is supported by the National Institutes of Health Clinical and Translational Science Award program, grants UL1 RR024153 and UL1TR000005. Data collection for the online survey was supported by the National Institutes of Health through Grant Number UL1TR000005.

Table of Contents

Table of Tables

Table of Figures

Overview

PROMIS® (Patient-Reported Outcomes Measurement Information System®) is a publicly available resource of a standardized, reliable, and efficient patient-reported outcome (PRO) measurement system for self-reported health domains (e.g., pain, fatigue, emotional distress, physical function, social function) that are relevant for both the general population and individuals with chronic illnesses [Cella et al. 2007; Cella et al. 2010]. PROMIS was constructed using Item Response Theory (IRT), which calibrates many items onto a single unidimensional construct for the domain relevant to that item. For example, the physical functioning domain is a unidimensional construct upon which 90 different items have been calibrated; at the high end of physical functioning are items like "able to run 100 yards," in the center are items like "able to climb stairs," and at the lower end are items like "able to get out of bed." All of the items together are called an "item bank" and any subset of items from the bank can be used to create a domain score. Because items are calibrated to the underlying construct, the domain score from any subset of items can be directly compared to a domain score from any other subset of items from the same domain item bank.

Health-related quality of life measures constructed using IRT represents a major advancement in health outcomes measurement. The banks are rigorously developed, generally avoid ceiling and floor effects, have known precision in domain score estimates, and allow flexible administration. Also leveraged in PROMIS is the conceptualization of a health domain as an underlying continuous construct.

At the time of the PROPr project, there were dozens of adult item banks available in PROMIS [see www.healthmeasures.net]. A researcher would choose a set of domains and collect and report scores for each domain. This could be a laborious process, however, since there was unfortunately no way to combine multiple domains into a single summary score of health.[1] A single summary score is advantageous for two reasons: it is desirable for ease of comparisons across groups and across time, and is necessary for applications like cost-effectiveness analyses and cost-utility analyses.

The goal of this project was to create a generic, societal, preference-based scoring system for PROMIS (adult) [Hanmer 2015]. *Generic* means that it should capture a core set of domains important to most people, regardless of their specific health conditions. *Societal* means that the score is meant to represent the aggregated preferences of the US society as a whole. *Preference-based* means that expressed preferences are used to give relative value to different levels of health, drawing heavily from multi-attribute utility theory (MAUT).

The creation of a MAUT measure has several steps. The first is to create a descriptive space for valuation. We use PROMIS item banks to create the descriptive space with the goal of *generic* measurement as described in "Selection of PROPr Domains." The second is to create descriptions for use in the valuation tasks as described in "Creation of Health-state Descriptions." In these descriptions,

---

[1] Having a single summary score of health was desirable, so two families of methods have been used to estimate a summary score of health using *discrete* item combinations in PROMIS. In particular, the PROMIS-29, a standardized profile instrument of 29 items, was used to predict the EQ-5D-3L and the HUI3 [Hays 2016, Revicki 2009]. A DCE valuation study was conducted to provide preference-based scores for the PROMIS-29 [Craig 2014]. This study is the first to use the *continuous* domain constructs from PROMIS item banks.

we explicitly link value to the underlying health domain construct. The third is to obtain valuations of the descriptions. We used a nationally representative sample with the goal of *societal* preferences. We used standard-gamble as our method of valuation, as this has the strongest theoretical underpinnings from MAUT. Both the sample and valuation methods are described in “Valuation Survey.” The fourth step is to combine these valuations into a scoring function, as described in “Preference-Based Scoring System Estimation.”

At the time of writing this report, PROPr has been validated in cross-sectional analyses. The PROPr score is correlated to legacy measures such as the EQ-5D-5L, HUI2, and HUI3. The PROPr score also acts in expected ways by age, sex, and presence of chronic conditions. These findings are discussed in “Preference-Based Scoring System Validity” and “Combining PROPr Scores with Legacy Measure Scores.”

Selection of PROPr Domains

Developing preference-based scores requires that representatives from the population make explicit trade-offs between different health states. Because these tradeoffs are cognitively complex, we sought to limit the number of domains while maintaining comprehensiveness. At the time of domain construct, there were 37 adult domains available in PROMIS.

For explicit trade-offs to make sense, domains must be structurally independent. Structural independence means that the range of possible outcomes on domain A is potentially independent of the outcome on domain B (and vice versa). For example, physical function and depression are structurally independent if one can imagine an individual with good levels of physical function and very severe depression as well as an individual who is not depressed with very poor physical function. Domains can be structurally independent even if they are highly correlated (e.g., depressed individuals tend to have poor physical functioning).

PROPr is intended to be a generic preference-based measure of health-related quality of life. With this perspective in mind, domain selection was completed in four steps:

Step 1: Modified Delphi reduction of available domains

A set of 9 experts in PROMIS, utility measurement, and cost-effectiveness analysis completed three rounds in an online Delphi procedure to reduce the number of domains assessed in steps 2 and 3. The goal was to remove domains which should not be in a generic health measure (e.g., alcohol use) and to choose a single domain from sets where there was substantial redundancy (e.g., one of three social roles domains was chosen). Ten domains were retained:

1. Cognitive Function – Abilities
2. Anxiety
3. Depression
4. Fatigue
5. Pain Intensity
6. Pain Interference
7. Physical Function
8. Ability to Participate in Social Roles and Activities (Social Roles)
9. Sleep Disturbance
10. Sexual Function

Step 2: Structural independence in a community sample

MAUT assumes that the domains in a descriptive space are structurally independent of each other, meaning that the range of possible responses on domain 1 is not constrained by the responses on domain 2. For example, most people think depression and physical functioning are structurally independent: 1) the degree of a person’s depression does not necessarily constrain how much physical functioning they might have and 2) the degree of a person’s physical functioning does not necessarily constrain how much depression they might have. This independence is conceptual and not epidemiologically observed (e.g., people with very bad depression are more likely to have poor physical functioning, but they don't have to have poor physical functioning). Structural independence refers to what is possible, rather than what is probable.

We recruited 50 community dwelling adults from the University of Pittsburgh's Clinical and Translational Science Institute's research registry website to evaluate if pairs of the health-related domains were structurally independent. Participants were required to be 18 years old or older and speak English; there were no exclusion criteria. The sample was 60% female with a mean age of 44 (range 22-70). Fifty-two percent of the sample were White, 32% were Black, and 7% were Other race. They gave self-rated health reports of: Excellent (27%), Very Good (41%), Good (29%), Fair (4%), and Poor (0%).

Participants were involved in face-to-face interviews during which they used a paired comparison task to evaluate 20 randomly assigned pairs of the health domains selected in Step 1. The research assistant explained that we were asking if the combinations were possible but not necessarily probable (i.e., could the respondent imagine such a combination ever happening). Participants were given the name and description of each domain on a piece of paper. The content of these descriptions were based on the definitions for the different item banks in PROMIS® (Cella et al., 2010; Riley et al., 2010).

Table 1 includes the proportion of respondents who reported that a particular combination of health-related domains was structurally independent. Most pairs were evaluated 22 times (range 20-27). The proportion of pairs reported to be structurally independent across all comparisons was 0.78.

Domain	Applied Cognition	Anxiety	Depression	Fatigue	Pain Interference	Pain Intensity	Physical Function	Social Roles	Sexual Function
Applied Cognition	-	-	-	-	-	-	-	-	-
Anxiety	0.68	-	-	-	-	-	-	-	-
Depression	0.68	0.81	-	-	-	-	-	-	-
Fatigue	0.68	0.91	0.85	-	-	-	-	-	-
Pain Interference	0.90	0.91	0.76	0.86	-	-	-	-	-
Pain Intensity	0.68	0.68	0.80	0.88	**0.61**	-	-	-	-
Physical Function	0.91	0.74	0.95	0.68	0.67	0.95	-	-	-
Social Roles	0.81	0.79	0.71	0.72	0.81	**0.55**	0.87	-	-
Sexual Function	0.91	0.91	0.73	**0.64**	0.77	**0.52**	0.86	0.95	-
Sleep Disturbance	0.68	**0.50**	0.86	0.90	0.76	0.77	0.77	0.86	0.74

Table 1: Respondents' evaluation of structural independence of health-related domains.

Step 3: Domain importance in a community sample

The same participants in step 2 were given cards with each of the 10 health domains, and asked to do several things: they were to remove a health domain if they did not think it important for overall quality of life; and they were asked to rank, order, and rate the relative importance of the remaining domains on a 0 to 100 visual analog scale (VAS), with the most important health-related domain placed at 100 and the least important domain at 0. We created a simple rank of the domains from a participant's VAS scores.

One participant, while able to complete step 2, was unable to comprehend and complete the task for step 3, so 49 participants were included in this analysis. One respondent removed Fatigue, Pain Intensity, Pain Interference, and Sexual Function, and three other respondents removed Sexual Function.

Table 2 lists the health domains, their mean VAS score, and mean rank.

Domain	N	VAS score, mean(sd)	Rank, mean (sd)
Physical Function	49	75 (30)	3.3 (2.5)
Applied Cognition	49	75 (30)	3.3 (2.7)
Pain Intensity	48	68 (32)	3.6 (2.4)
Sleep Disturbance	49	67 (27)	4.0 (2.5)
Depression	49	60 (33)	4.2 (2.8)
Pain Interference	48	55 (36)	4.6 (2.7)
Anxiety	49	50 (37)	5.1 (3.1)
Social Roles	49	53 (34)	5.7 (2.8)
Fatigue	48	48 (35)	5.9 (3.0)
Sexual Function	45	28 (34)	7.4 (2.7)

Table 2: Health Domains, Mean VAS score, and mean rank of each domain

Step 4: Merge

We used a combination of information from steps 2 and 3 to create the final set of domains for the health-state descriptive system for the PROPr score.

- First, we selected a core set of domains considered essential for face validity: physical functioning, depression, and one of the pain domains.
- Second, we used information from the community sample to determine which domains could be added to this core set while maintaining structural independence.
- Third, we used information about the health-related domain importance from the community sample to determine how best to include the most important domains from the participant's perspective.

The final set of domains used in PROPr development are:

- Cognitive Function – Abilities Subset v2.0 (Cognition)
- Emotional Distress – Depression v1.0 (Depression)

- Fatigue v1.0 (Fatigue)
- Pain Interference v1.1 (Pain)
- Physical Function v1.2 (Physical Function)
- Ability to Participate in Social Roles and Activities v2.0 (Social Roles)
- Sleep Disturbance v1.0 (Sleep)

Other versions of these domains give comparable scores to the domain versions used in PROPr development. As long as a domain version is considered comparable because of IRT co-calibration, it can be used to calculate a PROPr score. As of September 2017, comparable domains include:

- Cognition can also be measured with Cognitive Function v2.0.
- Pain can also be measured with Pain Interference v1.0.
- Physical Function can also be measured with Physical Function v1.0, v1.1, and v2.0.

For further details on domain selection, see Hanmer et al 2017.

Creation of Health-state Descriptions

PROMIS item banks have been created using item response theory. Many items and their responses are calibrated onto a unidimensional construct (such as depression or pain) called theta. Each item response has information about where a respondent is on theta. This information can be expressed in item characteristic curves (such as Figure 1). Theta is constructed such that 0 is the population mean with a standard deviation of 1. While theta scores are hypothetically unbounded, nearly all possible PROMIS scores are between -4 and +4.

We tested a variety of different methods to present a single item bank for valuation. The best method we found was to select two items to represent the item bank.

First, we had to simplify to the Item Information into something easily digestible. This was done by illustrating the theta score that each response would give if it was given in isolation as show in Figure 1.

Creating Point Estimates

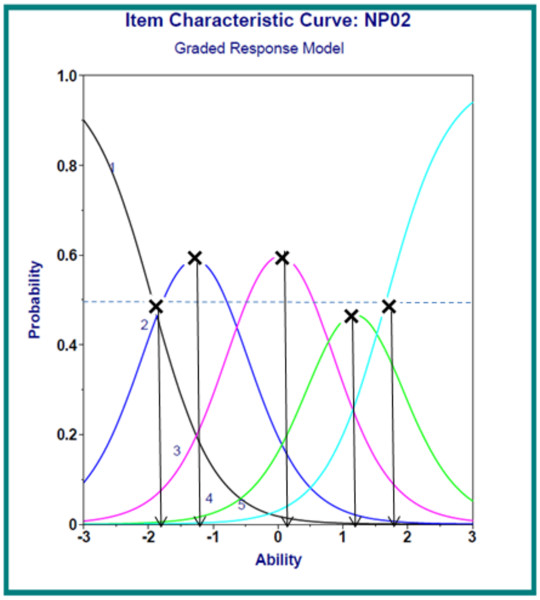


Figure 1: Creating point estimates using item characteristic curves

Then we selected the five items each of which captured the highest and lowest ends of the range. The Depression item bank is illustrated below. On the Y-axis is the item number. On the X-axis is the theta estimate each of the item responses would give in isolation. The lowest and highest theta estimates are highlighted:

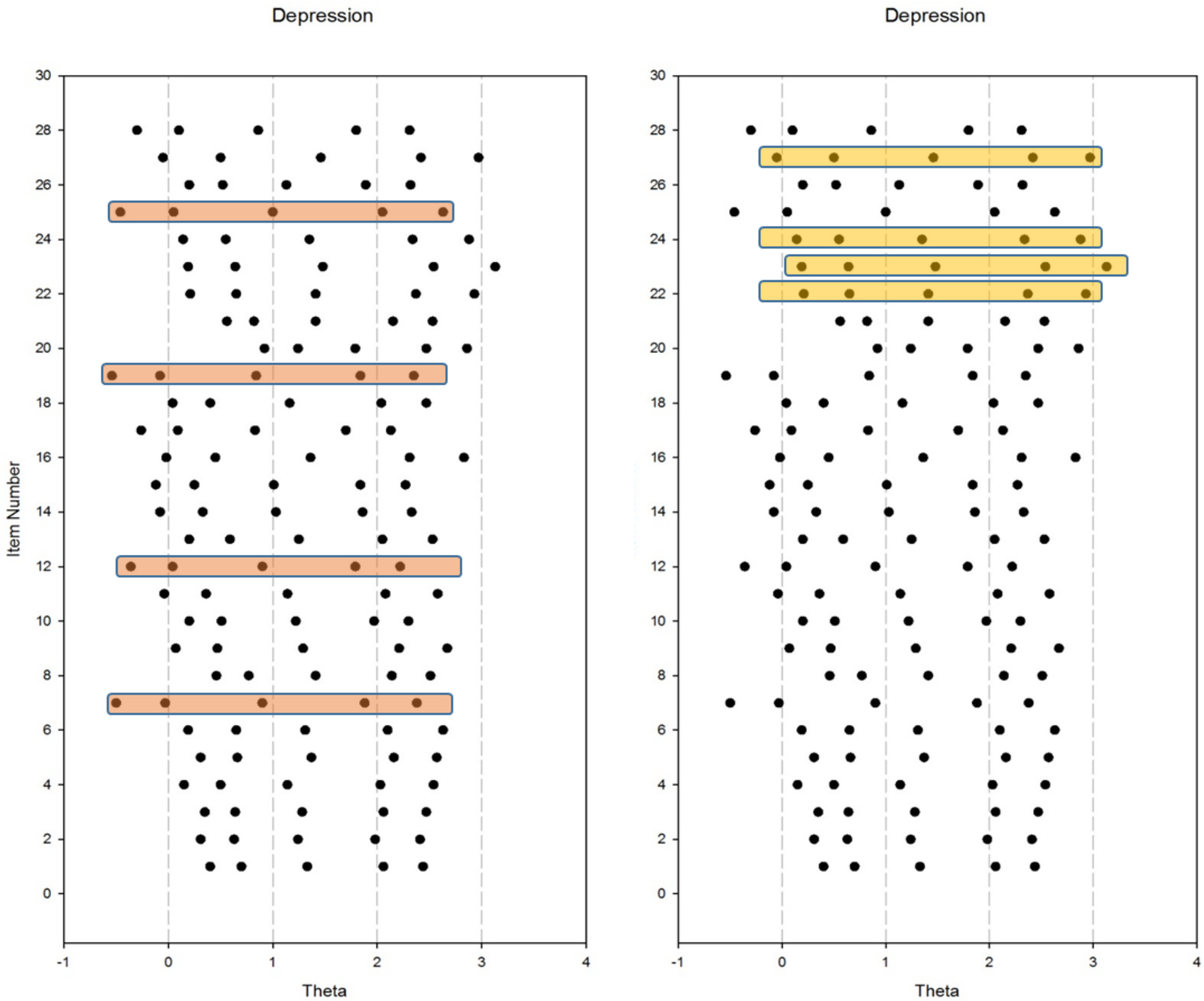


Figure 2: Example of selecting two items from an item bank

We then had experts evaluate the content of the questions.

Cognition

Item Bank: PROMIS Bank v2.0 – Cognitive Function Abilities Subset
Notes: Wanted to capture both attention and memory

Items selected from highest and lowest options:

I have been able to concentrate. . . PC6	Not at all	A little bit	Somewhat	Quite a bit	Very much
I have been able to remember to do things, like take medicine or buy something I needed . . . PC27	Not at all	A little bit	Somewhat	Quite a bit	Very much

Depression

Item Bank: PROMIS Item Bank v. 1.0 – Emotional Distress – Depression
Notes: Wanted to capture both sadness and apathy

Items selected from highest and lowest options:

I felt unhappy . . . EDDEP36	Never	Rarely	Sometimes	Often	Always
I felt that nothing was interesting . . . EDDEP45	Never	Rarely	Sometimes	Often	Always

Fatigue

Item Bank: PROMIS Item Bank v. 1.0 – Fatigue
Notes: N/A

Items selected from highest and lowest options:

How often were you too tired to take a bath or shower? . . . FATIMP21	Always	Often	Sometimes	Rarely	Never
How often did you feel tired? FATEXP20	Always	Often	Sometimes	Rarely	Never

Pain

Item Bank: PROMIS v1.1 Pain Interference
Notes: Difficult to find items which are independent of other domains

Items selected from highest and lowest options:

How often was your pain so severe you could think of nothing else? . . . PAININ29	Always	Often	Sometimes	Rarely	Never
How often was pain distressing to you?. . . PAININ24	Always	Often	Sometimes	Rarely	Never

Physical Function

Item Bank: PROMIS Bank v.1.2 – Physical Function
Notes: Wanted to capture both dexterity and mobility

Items selected from highest and lowest options:

Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .A16	Unable to do	With much difficulty	With some difficulty	With a little difficulty	Without any difficulty
Are you able to run 100 yards (100 m)? . . . pfc13	Unable to do	With much difficulty	With some difficulty	With a little difficulty	Without any difficulty

Sleep Disturbance

Item Bank: PROMIS Item Bank v. 1.0 – Sleep Disturbance
Notes: N/A

Items selected from highest and lowest options:

I got enough sleep . . .110	Never	Rarely	Sometimes	Often	Always
I woke up too early and could not fall back to sleep . . .50	Always	Often	Sometimes	Rarely	Never

Social Roles

Item Bank: PROMIS v2.0 Ability to Participate in Social Roles
Notes: Wanted to capture both discretionary and instrumental activities

Items selected from highest and lowest options:

I have trouble taking care of my regular personal responsibilities . . . SRPPER31_CaPS	Always	Usually	Sometimes	Rarely	Never
I have trouble participating in recreational activities with others. . . SRPPER04_CaPS	Always	Usually	Sometimes	Rarely	Never

The entire health-state description is:

Cognition	I have been able to concentrate. . .	Not at all	A little bit	Somewhat	Quite a bit	Very much
	I have been able to remember to do things, like take medicine or buy something I needed . . .	Not at all	A little bit	Somewhat	Quite a bit	Very much
Depre.	I felt unhappy . . .	Always	Often	Sometimes	Rarely	Never
	I felt that nothing was interesting . . .	Always	Often	Sometimes	Rarely	Never
Fatigu	How often were you too tired to take a bath or shower? . . .	Always	Often	Sometimes	Rarely	Never
	How often did you feel tired?	Always	Often	Sometimes	Rarely	Never
Pain	How often was your pain so severe you could think of nothing else? . . .	Always	Often	Sometimes	Rarely	Never
	How often was pain distressing to you?. . .	Always	Often	Sometimes	Rarely	Never
Physical	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	Without any difficulty
	Are you able to run 100 yards (100 m)? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	Without any difficulty
Sleep	I got enough sleep . . .	Never	Rarely	Sometimes	Often	Always
	I woke up too early and could not fall back to sleep . . .	Always	Often	Sometimes	Rarely	Never
Social Roles	I have trouble taking care of my regular personal responsibilities . . .	Always	Usually	Sometimes	Rarely	Never
	I have trouble participating in recreational activities with others. . .	Always	Usually	Sometimes	Rarely	Never

Table 3: Entire PROPr state description

For any individual domain, there are two items with five responses each. This allows for a set of 25 possible responses. However, if we look at these responses based on the area of theta which they are most associated, we can reduce this set of 25 possible responses to 9 likely responses.

This concept is illustrated in Figure 3. In the illustration, there are two items calibrated to the same item bank. Responses to the items are associated with particular areas of theta as shown in the colored bars below the item characteristic curves:

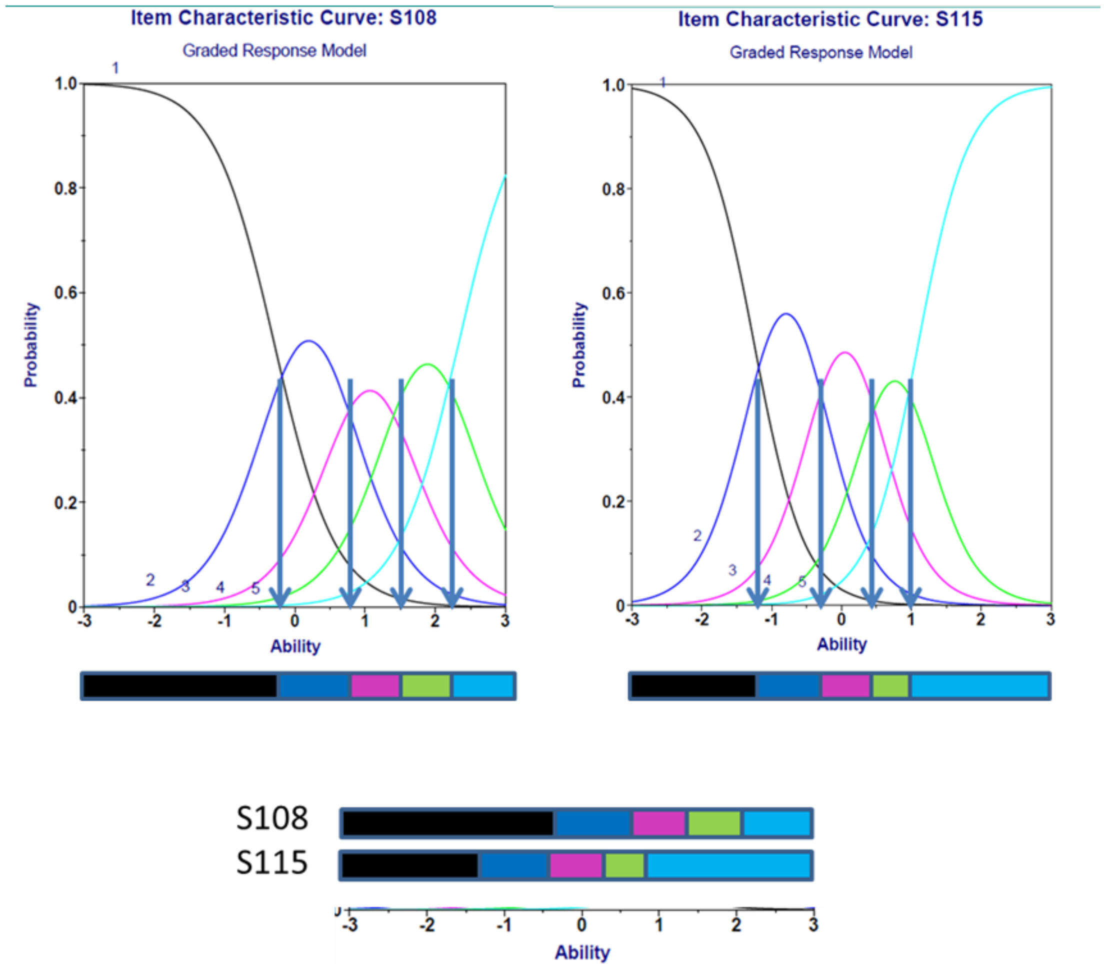


Figure 3: Illustration of combining two items

These colored bars can be overlaid to see what combination of colors is most likely for any point along theta. For instance, at -2, the most likely colors are S108=black and S115=black. For +1, the most likely colors are S108= magenta and S115= light blue. There are a total of 9 possible logical color combinations for any two items.

The cut points for each item are below. The combination of responses presented for valuation is included in Appendix 3.

Cognition

I have been able to concentrate. . .	-1.86	-0.80	-0.66	0.512
I have been able to remember to do things, like take medicine or buy something I needed . . .	-1.91	-1.48	-0.62	0.180

Depression

I felt unhappy . .	-0.51	0.34	1.35	2.33
I felt that nothing was interesting . . .	0.23	0.87	1.84	2.84

Fatigue

How often were you too tired to take a bath or shower? . . .	0.63	1.04	2.01	3.24
How often did you feel tired?	-1.61	-0.48	0.74	1.77

Pain

How often was your pain so severe you could think of nothing else? . . .	0.765	1.074	1.773	2.906
How often was pain distressing to you?. . .	0.054	0.597	1.312	2.037

Physical Function

Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	-3.04	-2.58	-1.92	-1.31
Are you able to run 100 yards (100 m)? . . .	-0.34	-0.71	-0.29	0.41

Sleep Disturbance

I got enough sleep . . .	-0.53	0.27	1.7	2.97
I woke up too early and could not fall back to sleep . . .	-1.51	-0.07	0.73	1.7

Social Roles

I have trouble taking care of my regular personal responsibilities . . .	-1.84	-1.57	-0.628	0.203
I have trouble participating in recreational activities with others. . .	-1.31	-0.82	-0.016	0.651

Theta Point Estimates for Each Domain

Each domain therefore has 9 health states for valuation (except Sleep Disturbance as both items have a cutpoint at 1.7). The theta scores for each health state are in Table 4.

	Best								Worst
Cognition	1.124 (0.689)	0.52 (0.591)	-0.002 (0.524)	-0.367 (0.55)	-0.649 (0.521)	-0.902 (0.534)	-1.239 (0.541)	-1.565 (0.572)	-2.052 (0.646)
Depression	-1.082 (0.617)	-0.264 (0.426)	0.151 (0.385)	0.596 (0.384)	0.913 (0.397)	1.388 (0.398)	1.742 (0.406)	2.245 (0.43)	2.703 (0.479)
Fatigue	-1.648 (0.611)	-0.818 (0.506)	-0.094 (0.487)	0.303 (0.447)	0.87 (0.436)	1.124 (0.441)	1.688 (0.48)	2.053 (0.508)	2.423 (0.61)
Pain	-0.773 (0.67)	0.1 (0.404)	0.462 (0.413)	0.827 (0.331)	1.072 (0.349)	1.407 (0.345)	1.724 (0.368)	2.169 (0.404)	2.725 (0.492)
Physical Function	0.966 (0.666)	0.16 (0.474)	-0.211 (0.451)	-0.443 (0.443)	-0.787 (0.49)	-1.377 (0.417)	-1.784 (0.465)	-2.174 (0.526)	-2.575 (0.622)
Sleep Disturbance	-1.535 (0.687)	-0.775 (0.615)	-0.459 (0.592)	0.093 (0.568)	0.335 (0.569)	0.82 (0.584)	1.659 (0.659)	1.934 (0.699)	
Social Roles	1.221 (0.576)	0.494 (0.372)	0.083 (0.341)	-0.276 (0.337)	-0.618 (0.347)	-0.955 (0.328)	-1.293 (0.378)	-1.634 (0.363)	-2.088 (0.493)

Table 4: Theta point estimates and standard deviation of the estimate for each health state in each domain in PROPr

To add some context about this range of scores, Figure 4 illustrates the theta scores for each health state. The dashed lines are the 5th and 95th percentile in the US population as measured by the calibration sample used to make the item parameters:

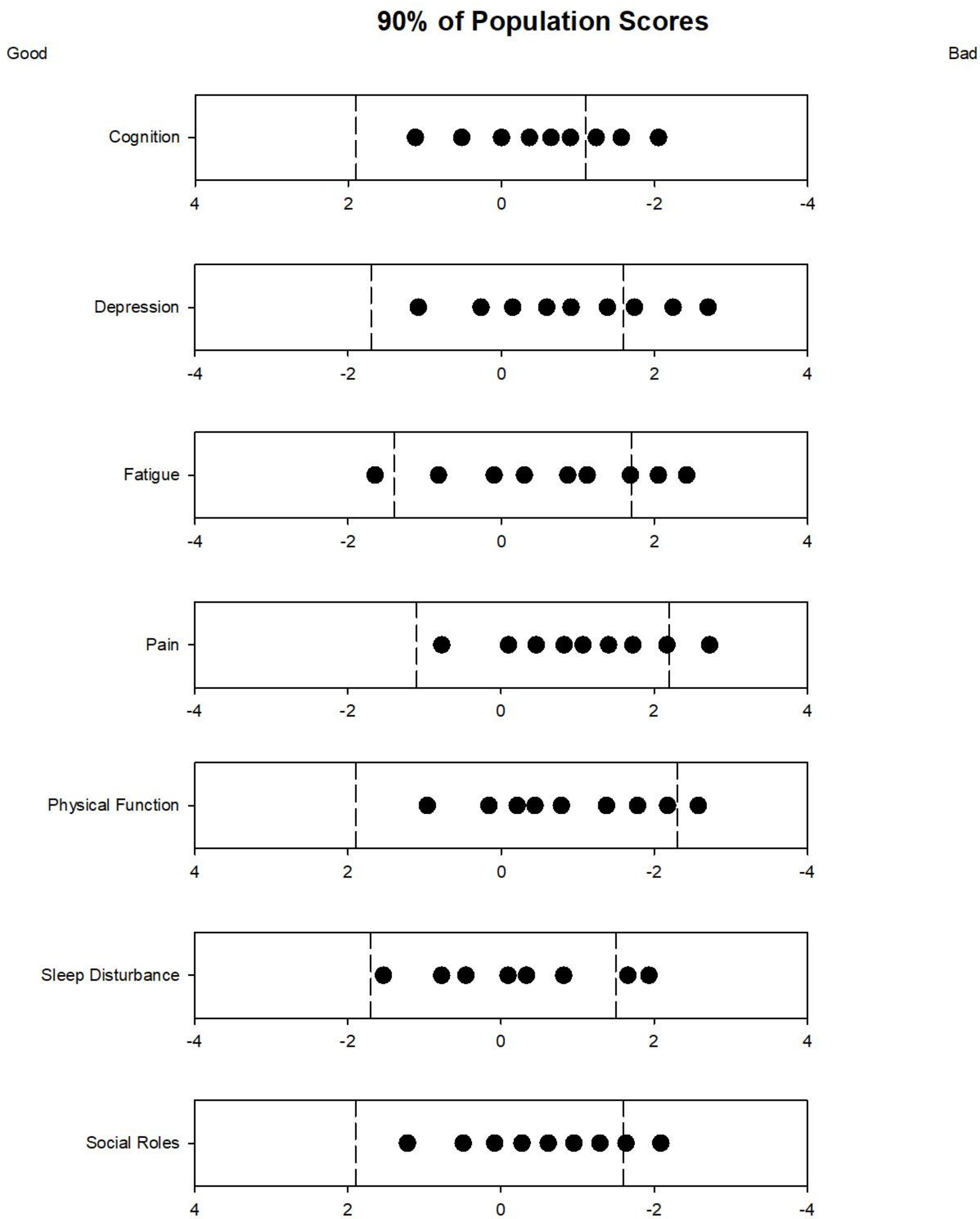


Figure 4: PROPr point estimates and range of general population scores

Valuation Survey

Definition of Terms

Full health – the health description with all domains at their best description:

Cognition	I have been able to concentrate. . .	Not at all	A little bit	Somewhat	Quite a bit	**Very much**
	I have been able to remember to do things, like take medicine or buy something I needed . . .	Not at all	A little bit	Somewhat	Quite a bit	**Very much**
Depression	I felt unhappy . . .	Always	Often	Sometimes	Rarely	**Never**
	I felt that nothing was interesting . . .	Always	Often	Sometimes	Rarely	**Never**
Fatigue	How often were you too tired to take a bath or shower? . . .	Always	Often	Sometimes	Rarely	**Never**
	How often did you feel tired?	Always	Often	Sometimes	Rarely	**Never**
Pain	How often was your pain so severe you could think of nothing else? . . .	Always	Often	Sometimes	Rarely	**Never**
	How often was pain distressing to you?. . .	Always	Often	Sometimes	Rarely	**Never**
Physical Function	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	**Without any difficulty**
	Are you able to run 100 yards (100 m)? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	**Without any difficulty**
Sleep	I got enough sleep . . .	Never	Rarely	Sometimes	Often	**Always**
	I woke up too early and could not fall back to sleep . . .	Always	Often	Sometimes	Rarely	**Never**
Social Roles	I have trouble taking care of my regular personal responsibilities . . .	Always	Usually	Sometimes	Rarely	**Never**

Figure 5: Full health

Corner State – the health description with all domains at their best description except one domain which is at its worst description. This is the Depression corner state:

Cognition	I have been able to concentrate. . .	Not at all	A little bit	Somewhat	Quite a bit	**Very much**
	I have been able to remember to do things, like take medicine or buy something I needed . . .	Not at all	A little bit	Somewhat	Quite a bit	**Very much**
Depression	I felt unhappy . . .	**Always**	Often	Sometimes	Rarely	Never
	I felt that nothing was interesting . . .	**Always**	Often	Sometimes	Rarely	Never
Fatigue	How often were you too tired to take a bath or shower? . . .	Always	Often	Sometimes	Rarely	**Never**
	How often did you feel tired?	Always	Often	Sometimes	Rarely	**Never**
Pain	How often was your pain so severe you could think of nothing else? . . .	Always	Often	Sometimes	Rarely	**Never**
	How often was pain distressing to you?. . .	Always	Often	Sometimes	Rarely	**Never**
Physical Function	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	**Without any difficulty**
	Are you able to run 100 yards (100 m)? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	**Without any difficulty**
Sleep	I got enough sleep . . .	Never	Rarely	Sometimes	Often	**Always**
	I woke up too early and could not fall back to sleep . . .	Always	Often	Sometimes	Rarely	**Never**
Social Roles	I have trouble taking care of my regular personal responsibilities . . .	Always	Usually	Sometimes	Rarely	**Never**

Figure 6: Example corner state

All-worst health state – the health state description with all domains at their worst description:

Cognition	I have been able to concentrate. . .	**Not at all**	A little bit	Somewhat	Quite a bit	Very much
	I have been able to remember to do things, like take medicine or buy something I needed . . .	**Not at all**	A little bit	Somewhat	Quite a bit	Very much
Depression	I felt unhappy . . .	**Always**	Often	Sometimes	Rarely	Never
	I felt that nothing was interesting . . .	**Always**	Often	Sometimes	Rarely	Never
Fatigue	How often were you too tired to take a bath or shower? . . .	**Always**	Often	Sometimes	Rarely	Never
	How often did you feel tired?	**Always**	Often	Sometimes	Rarely	Never
Pain	How often was your pain so severe you could think of nothing else? . . .	**Always**	Often	Sometimes	Rarely	Never
	How often was pain distressing to you?. . .	**Always**	Often	Sometimes	Rarely	Never
Physical Function	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	**Unable to do**	With much difficulty	With some difficulty	With a little difficulty	Without any difficulty
	Are you able to run 100 yards (100 m)? . . .	**Unable to do**	With much difficulty	With some difficulty	With a little difficulty	Without any difficulty
Sleep	I got enough sleep . . .	**Never**	Rarely	Sometimes	Often	Always
	I woke up too early and could not fall back to sleep . . .	**Always**	Often	Sometimes	Rarely	Never
Social Roles	I have trouble taking care of my regular personal responsibilities . . .	**Always**	Usually	Sometimes	Rarely	Never

Figure 7: All-worst health state

Marker State – a health state that has several domains which are not at the best or worst description. Three marker states were constructed which nest by severity:

Cognition	I have been able to concentrate. . .	Not at all	A little bit	Somewhat	Quite a bit	**Very much**
	I have been able to remember to do things, like take medicine or buy something I needed . . .	Not at all	A little bit	Somewhat	Quite a bit	**Very much**
Depression	I felt unhappy . . .	Always	Often	Sometimes	Rarely	**Never**
	I felt that nothing was interesting . . .	Always	Often	Sometimes	Rarely	**Never**
Fatigue	How often were you too tired to take a bath or shower? . . .	Always	Often	Sometimes	Rarely	**Never**
	How often did you feel tired?	Always	Often	Sometimes	Rarely	**Never**
Pain	How often was your pain so severe you could think of nothing else? . . .	Always	Often	**Sometimes**	Rarely	Never
	How often was pain distressing to you?. . .	Always	Often	**Sometimes**	Rarely	Never
Physical Function	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	**Without any difficulty**
	Are you able to run 100 yards (100 m)? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	**Without any difficulty**
Sleep	I got enough sleep . . .	Never	Rarely	Sometimes	**Often**	Always
	I woke up too early and could not fall back to sleep . . .	Always	Often	Sometimes	**Rarely**	Never
Social Roles	I have trouble taking care of my regular personal responsibilities . . .	Always	Usually	Sometimes	Rarely	**Never**
	I have trouble participating in recreational activities with others. . .	Always	Usually	Sometimes	Rarely	**Never**

Figure 8: Marker state A

Cognition	I have been able to concentrate. . .	Not at all	A little bit	Somewhat	Quite a bit	**Very much**
	I have been able to remember to do things, like take medicine or buy something I needed . . .	Not at all	A little bit	Somewhat	Quite a bit	**Very much**
Depression	I felt unhappy . . .	Always	Often	Sometimes	**Rarely**	Never
	I felt that nothing was interesting . . .	Always	Often	Sometimes	**Rarely**	Never
Fatigue	How often were you too tired to take a bath or shower? . . .	Always	Often	Sometimes	Rarely	**Never**
	How often did you feel tired?	Always	Often	Sometimes	Rarely	**Never**
Pain	How often was your pain so severe you could think of nothing else? . . .	Always	Often	**Sometimes**	Rarely	Never
	How often was pain distressing to you?. . .	Always	Often	**Sometimes**	Rarely	Never
Physical Function	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	**Without any difficulty**
	Are you able to run 100 yards (100 m)? . . .	Unable to do	**With much difficulty**	With some difficulty	With a little difficulty	Without any difficulty
Sleep	I got enough sleep . . .	Never	Rarely	Sometimes	**Often**	Always
	I woke up too early and could not fall back to sleep . . .	Always	Often	Sometimes	**Rarely**	Never
Social Roles	I have trouble taking care of my regular personal responsibilities . . .	Always	Usually	Sometimes	Rarely	**Never**
	I have trouble participating in recreational activities with others. . .	Always	Usually	Sometimes	Rarely	**Never**

Figure 9: Marker state B

Cognition	I have been able to concentrate. . .	Not at all	A little bit	**Somewhat**	Quite a bit	Very much
	I have been able to remember to do things, like take medicine or buy something I needed . . .	Not at all	A little bit	**Somewhat**	Quite a bit	Very much
Depression	I felt unhappy . . .	Always	Often	Sometimes	**Rarely**	Never
	I felt that nothing was interesting . . .	Always	Often	Sometimes	**Rarely**	Never
Fatigue	How often were you too tired to take a bath or shower? . . .	Always	Often	Sometimes	Rarely	**Never**
	How often did you feel tired?	Always	Often	Sometimes	Rarely	**Never**
Pain	How often was your pain so severe you could think of nothing else? . . .	**Always**	Often	Sometimes	Rarely	Never
	How often was pain distressing to you?. . .	**Always**	Often	Sometimes	Rarely	Never
Physical Function	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	**Without any difficulty**
	Are you able to run 100 yards (100 m)? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	**Without any difficulty**
Sleep	I got enough sleep . . .	Never	Rarely	Sometimes	**Often**	Always
	I woke up too early and could not fall back to sleep . . .	Always	Often	Sometimes	**Rarely**	Never
Social Roles	I have trouble taking care of my regular personal responsibilities . . .	Always	Usually	Sometimes	Rarely	**Never**
	I have trouble participating in recreational activities with others. . .	Always	Usually	Sometimes	Rarely	**Never**

Figure 10: Marker state C

Dead – labelled as dead (a health state) not as death (a process).

Survey Outline

1. Consent
2. Demographic Questions
3. Single question of self-rated health:
   "In general, would you say your health is: Excellent, very good, good, fair, or poor"
4. Randomize 1:1:1:1 with Module 1: Module 2: Module 3: Module 4 as below
5. PROMIS 29 + cognition short form (4 question version)
6. Self-assessed additional life expectancy
7. Random assignment to 1 domain and 2 other corner/marker states.
   a. Domain valuation task with "full health" as the best health state and the domain corner state as the worst health state. The respondent valued the remaining 7 domain levels first by visual analog scale (VAS) and then by standard gamble (SG). Domain levels were randomized.
   b. Selection of dead or "all-worst health state" as the worst health state.
      i. If dead is the worst health state: Corner/Marker state valuation task with full health as the best health state and the selected dead as the worst health state. The respondent valued the 3 corner/marker states and remaining all-worst health state first by VAS and then by SG. The four states were randomized.
      ii. If all-worst health state is the worst health state: Corner/Marker state valuation task with full health as the best health state and the selected dead as the worst

health state. The respondent valued the 3 corner/marker states and remaining all-worst health state first by VAS and then by SG. The four states were randomized.

8. Task engagement questions
9. Remaining modules – randomized

Modular Pieces:

1. PROMIS-Global
2. EQ-5D-5L with VAS
3. Health Utilities Index Mark 2 and Mark 3
4. Chronic Health Conditions List
5. Numeracy
6. Experience with disability
7. Distributional preferences - social (1 randomly assign one from set of 5)
8. Distributional preferences - personal (1 randomly assign one from set of 5)

Response Rates and Drop Out

2026 invitations

1779 completed the consent form (87.8% of those invited)

27 dropped out during demographics
304 dropped out between demographics and start of valuation
242 dropped out during VAS
37 dropped out during SG
5 dropped out after SG

1164 completes (57.5% of those invited, 65.4% of those who completed the consent)

Survey Sampling

The survey was administered by Research Now which maintains a panel of pre-validated individuals, primarily for market research. Once enrolled in the panel, participants are invited to complete surveys for incentives that are redeemable for gift cards and points programs. Incentives are based on the amount of time estimated to complete a survey and are only given at survey completion.

Survey Demographics

The survey was intended to be representative of the US population by age, sex, race, ethnicity, education, and income with 1000 respondents. Filling some cells, particularly low education, was more difficult than expected. To fill all cells to their quota, 1164 responses were collected. Data were collected in spring 2016.

The survey was estimated to take 30 to 35 minutes. Given that the survey designers could not complete it in less than 15 minutes, we decided to exclude participants who took less than 15 minutes. We have explored a wide variety of other exclusion criteria with minimal change in preference based scoring system results.

Gender	Quota	Total	Final
Female	510	614	532
Male	490	547	451
Age	Quota	Total	Final
18 – 24	130	135	95
25 – 34	170	214	157
35 – 44	170	174	142
45 – 54	190	195	173
55 – 64	160	193	172
65 – 74	90	128	126
75 – 84	60	70	69
85 +	30	55	50
Hispanic	Quota	Total	Final
Yes	160	198	154
No	840	966	830
Race	Quota	Total	Final
White	720	878	758
African American	120	145	115
American Indian	10	12	10
Asian	50	64	44
Native Hawaiian	10	2	2
Other	60	37	35
Multiple Races	30	26	20
Education	Quota	Total	Final
Less than high school graduate	121	122	109
High school graduate or equivalent	244	271	249
Some college, no degree	183	223	221
Associate's degree	69	71	65
Bachelor's degree	157	200	216
Graduate or professional degree	96	142	124
Income	Quota	Total	Final
Less than $10,000	20	43	33
$10,000 to less than $15,000	40	41	37
$15,000 to less than $25,000	140	120	104
$25,000 to less than $35,000	170	184	156
$35,000 to less than $50,000	200	215	175
$50,000 to less than $65,000	150	191	166
$65,000 to less than $75,000	60	70	61
$75,000 to less than $100,000	100	129	108
$100,000 or more	120	171	144

Table 5: Online survey demographics

Preference-Based Scoring System Estimation

The PROPr score is estimated using multi-attribute utility theory (MAUT) (Keeney & Raiffa, 1976). A utility function can be built for a set of options using MAUT if each option can be decomposed using a finite number of common attributes, as is the case for PROPr. MAUT provides the forms these functions take, based on various relationships between the attributes, such as the structural independence conditions explained in the Domain Selection section.

In general, there are three main multi-attribute models: the linear additive model, the multiplicative model, and the multi-linear model. The linear additive model assumes that there are no preference interactions between the attributes; the multiplicative model allows for attributes to be either preference substitutes or complements (but not both); and the multi-linear form allows for more complicated interactions. The linear additive model makes the most restrictive assumptions—and, as we will see, is rejected by the data. The multiplicative model is, in fact, the simplest non-additive multi-attribute model, and has been shown to work well in previous multi-attribute health utility systems, such as the HUI:3 (Furlong et al., 1998). The multi-linear model, though attractive for the range of interactions it can capture, imposes a large data burden for estimation. Thus, based on previous work (Furlong et al., 1998), data was collected with the multiplicative model in mind. This design choice is checked using the data and a theorem from MAUT (Keeney & Raiffa, 1976, Appendix 6B; Torrance et al., 1996).

For more thorough detail on MAUT, in particular on its use in the health utility context, see Furlong et al. (1998).

Single-attribute Utility Functions

The standard gamble algorithm (Furlong et al., 1998; p. 64) allowed for valuations at each 0.05 (i.e., 0, 0.05, 0.10, 0.15, …, 0.95, 1.0). The algorithm also allowed respondents to have valuations >1.0 and <0 (e.g., the respondent chose the intermediate health state over a gamble where full health had a probability of 100%). Because we had no information about how far above 1.0 or below 0 these valuations would be, we rounded these values to 1.0 and 0, respectively.

As described previously in the Survey Outline, participants were randomly assigned to a domain and then valued multiple states within the domain. For each state, we trimmed 5% of the highest and lowest valuations ("10% trimming"), consistent with previous work (Feeny et al., 2002). For every domain, we then constructed a utility function—the so-called *single-attribute* utility function for that domain.

Previous single-attribute functions, such as those in the HUI:2 or HUI:3, are estimated over a discrete state space. In contrast, each PROMIS domain is a continuous construct (i.e., theta). That raises the question of how to estimate utilities for states between the levels of theta corresponding to the health state descriptions valued by the participants (Table 4). In addition, multi-attribute utility theory constrains the endpoints of each single-attribute function, requiring that a utility (disutility) of 1 (0) is assigned to full health, and a utility (disutility) of 0 (1) is assigned to the disutility corner state of the domain. Normative economic theory also requires that the function be monotonically increasing as functional capacity increases, so that one does not pay for treatments that *worsen* health. Thus, the single-attribute (dis)utility functions must be monotonic with fixed endpoints.

We explored several alternative specifications of the single-attribute functions: non-parametric models (i.e., smooth splines and kernel regression), as well as various parametric models (e.g., polynomial regression, varying the degree of the polynomial). Although the non-parametric models have desirable statistical properties (e.g., kernel regression will eventually recover the true regression function), in practice, their curvature sometimes violated the monotonicity assumption, or changed drastically over the range of theta where we have no data – changes that we cannot verify. They also lack parsimony. In contrast, while the parametric models are parsimonious, they were sometimes non-monotonic or exhibited similar changes in curvature over values of theta where there is no data.

As a result, we adopt the following procedure, which we believe strikes the right balance between producing unbiased estimates of the utilities, model parsimony, satisfying the theoretical requirements of MAUT, and avoiding functional forms that cannot currently be verified using out-of-sample data (or through cross-validation): we combine isotonic regression with linear interpolation. Isotonic regression estimates the sample mean values as the mean regression value – which are unbiased estimates of the mean – so long as those sample means are monotonic; when they are not, it calculates a weighted average of the two means and assigns that value, enforcing monotonicity. (There were only three instances of non-monotonicity across all domains.) Then, we connect each mean value with a line to estimate the utilities for theta values where we have no data, so that the curvature of the utility function in each gap of the data is the same.

The single-attribute functions, along with some of the alternatives described above, are plotted in Figure 11. In general, the isotonic regression combined with linear interpolation tracks the other models well, while satisfying all model requirements (e.g., monotonicity). Note that the functions are defined in terms of *disutility*, which is equal to 1 minus utility. The reason for this will be explained in the next subsection. The figure illustrates the link between a theta score for a given domain and the corresponding single-attribute (dis)utility function. Table 6 through Table 12 show the results of the isotonic regression combined with the linear interpolation, describing each point and line between consecutive points, for each domain.

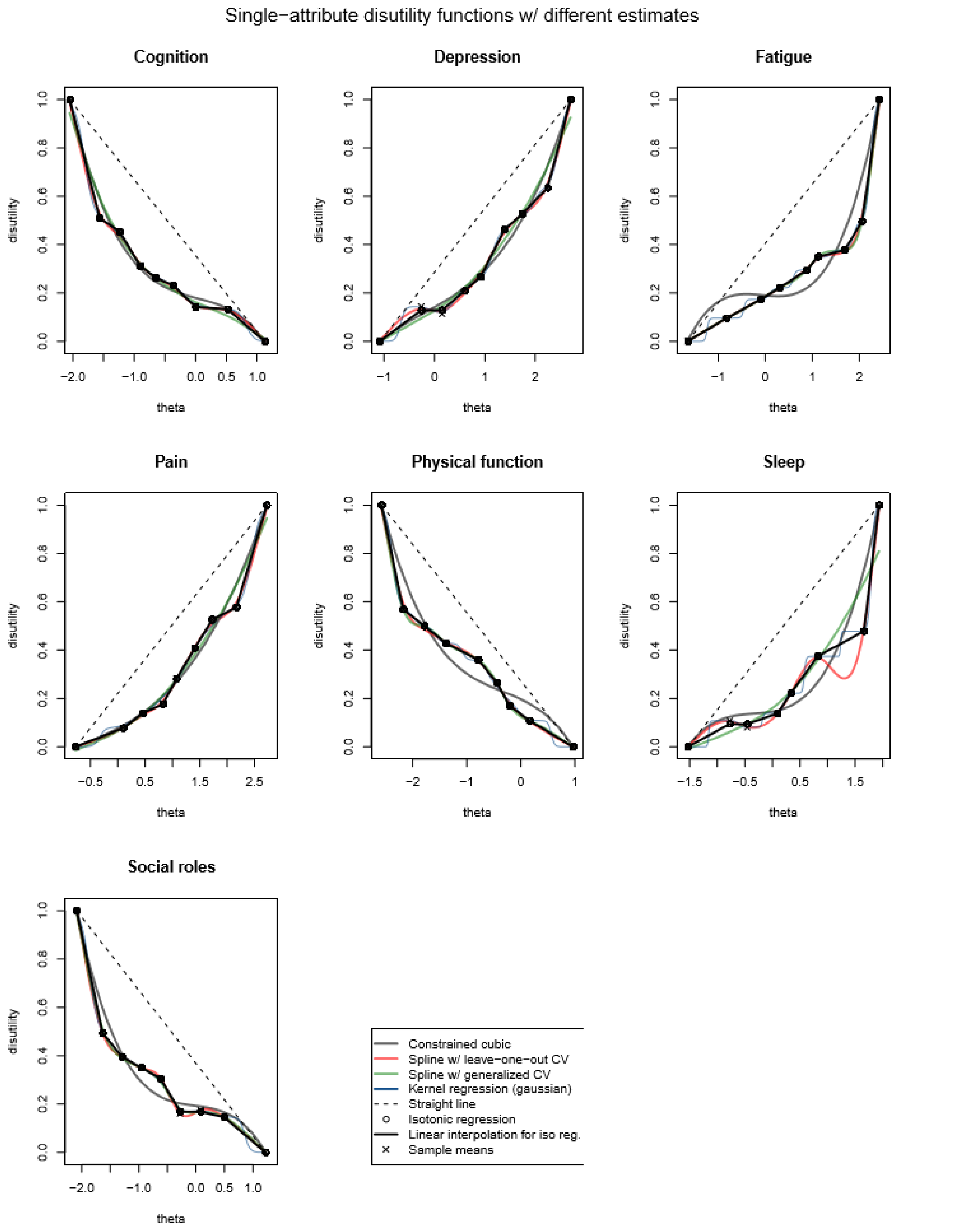


Figure 11: Single-attribute disutility functions for each domain. The linear interpolation with isotonic regression was ultimately selected. Note that a constrained *linear model (dotted line) would be wholly defined by the endpoint constraints, and thus not be a function of the data.*

Cognition Isotonic Regression Results			
Theta Range		Disutility = Intercept + slope(theta)	
Lower Bound	**Upper Bound**	**Intercept**	**Slope**
none	-2.052	1	0
-2.052	-1.565	-1.0617	-1.0047
-1.565	-1.239	0.2375	-0.1745
-1.239	-0.902	-0.0694	-0.4223
-0.902	-0.649	0.1357	-0.1949
-0.649	-0.367	0.192	-0.1082
-0.367	-0.002	0.1411	-0.2468
-0.002	0.52	0.1416	-0.0176
0.52	1.124	0.2464	-0.2192
1.124	none	0	0

Table 6: Cognition Isotonic Regression Results

Depression Isotonic Regression Results			
Theta Range		Disutility = Intercept + slope(theta)	
Lower Bound	**Upper Bound**	**Intercept**	**Slope**
none	-1.082	0	0
-1.082	-0.264	0.1701	0.1572
-0.264	0.151	0.1286	0
0.151	0.596	0.1015	0.1793
0.596	0.913	0.1001	0.1817
0.913	1.388	-0.1092	0.4109
1.388	1.742	0.1993	0.1887
1.742	2.245	0.1595	0.2115
2.245	2.703	-1.1577	0.7983
2.703	none	1	0

Table 7: Depression Isotonic Regression Results

Fatigue Isotonic Regression Results			
Theta Range		Disutility = Intercept + slope(theta)	
Lower Bound	**Upper Bound**	**Intercept**	**Slope**
none	-1.648	0	0
-1.648	-0.818	0.1898	0.1152
-0.818	-0.094	0.1837	0.1077
-0.094	0.303	0.1848	0.1189
0.303	0.87	0.1821	0.1277
0.87	1.124	0.1	0.222
1.124	1.688	0.2938	0.0496
1.688	2.053	-0.1681	0.3233
2.053	2.423	-2.3031	1.3632
2.423	none	1	0

Table 8: Fatigue Isotonic Regression Results

Pain Interference Isotonic Regression Results			
Theta Range		Disutility = Intercept + slope(theta)	
Lower Bound	**Upper Bound**	**Intercept**	**Slope**
none	-0.773	0	0
-0.773	0.1	0.0689	0.0891
0.1	0.462	0.0606	0.1721
0.462	0.827	0.0929	0.1022
0.827	1.072	-0.1733	0.4241
1.072	1.407	-0.1277	0.3815
1.407	1.724	-0.1089	0.3681
1.724	2.169	0.3243	0.1169
2.169	2.725	-1.0692	0.7594
2.725	none	1	0

Table 9: Pain Interference Isotonic Regression Results

Physical Function Isotonic Regression Results			
Theta Range		Disutility = Intercept + slope(theta)	
Lower Bound	**Upper Bound**	**Intercept**	**Slope**
none	-2.575	1	0
-2.575	-2.174	-1.7709	-1.0761
-2.174	-1.784	0.1867	-0.1756
-1.784	-1.377	0.1853	-0.1764
-1.377	-0.787	0.2683	-0.1161
-0.787	-0.443	0.1456	-0.2721
-0.443	-0.211	0.0853	-0.4082
-0.211	0.16	0.1356	-0.1695
0.16	0.966	0.13	-0.1346
0.966	none	0	0

Table 10: Physical Function Isotonic Regression Results

Sleep Disturbance Isotonic Regression Results			
Theta Range		Disutility = Intercept + slope(theta)	
Lower Bound	**Upper Bound**	**Intercept**	**Slope**
none	-1.535	0	0
-1.535	-0.775	0.1905	0.1241
-0.775	-0.459	0.0943	0
-0.459	0.093	0.1309	0.0797
0.093	0.335	0.1062	0.3455
0.335	0.82	0.1164	0.3148
0.82	1.659	0.2731	0.1238
1.659	1.934	-2.6676	1.8964
1.934	none	1	0

Table 11: Sleep Disturbance Isotonic Regression Results

Social Roles Isotonic Regression Results			
Theta Range		Disutility = Intercept + slope(theta)	
Lower Bound	Upper Bound	Intercept	Slope
none	-2.088	1	0
-2.088	-1.634	-1.3285	-1.1152
-1.634	-1.293	0.0241	-0.2874
-1.293	-0.955	0.2209	-0.1352
-0.955	-0.618	0.2239	-0.132
-0.618	-0.276	0.0576	-0.4012
-0.276	0.083	0.1683	0
0.083	0.494	0.1728	-0.054
0.494	1.221	0.2454	-0.201
1.221	none	0	0

Table 12: Social Roles Isotonic Regression Results

Disutility Corner States

The entire multi-attribute scoring function was built as a disutility function, because of the corner states. *Disutility corner states*—where every attribute is at its best save for the domain under consideration, which is at its worst description—are easier for participants to value than *utility corner states*, which have the complementary description (Torrance et al., 1996).

The values of the disutility corner states are an integral part of the multi-attribute scoring function. This can be seen in the definition of the multiplicative form of the function:

$$(1) \qquad u(\Theta) = 1 - \frac{\bar{u}(\Theta)}{\bar{u}(dead)},$$

where

$$\bar{u}(\Theta) = \frac{1}{c} \prod_{i=1}^{7} (1 + c \cdot c_i \cdot \bar{u}_i(\theta_i))$$

Here, $\Theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7)$, is the vector of PROMIS scores (i.e., the health status under consideration). The function $\bar{u}$ is the *disutility* multi-attribute function on the All-worst health state=1 to Full Health=0 scale (equation (1)), and so $\bar{u}(dead)$ is the disutility of dead on that scale. The constant c is the global interaction constant (in disutility terms), c_i is the mean disutility corner state disutility value for domain i, and $\bar{u}_i$ is the single-attribute disutility function for that domain. Reading equation (1) from the right to the left makes each step explicit: multiply the multi-attribute disutility function on the All-worst health state=1/Full Health=0 scale by $\frac{1}{\bar{u}(dead)}$ to turn it into the multi-attribute disutility function on dead-full health scale[2], and then subtract it from 1 to make it the utility function on the Dead=0 to Full Health=1 scale.

For a given domain, its disutility corner state was valued by those who valued that entire domain, as well as a random number of other participants who were assigned to value that domain's corner state as part of the random assignment of other states (see Survey Overview).

Calculation of the Global Scaling Constant

As can be seen in equation (1), the multi-attribute utility function also requires another constant (c), called the *global scaling constant*. This constant is the solution to the following equation:

$$\left(\prod_{i=1}^{7}(1 + c \cdot c_i)\right) - c - 1 = 0.$$

Most statistical software has the functionality to solve a single-variable polynomial of this form. For example, the following code, written in R, will find the roots of the polynomial.

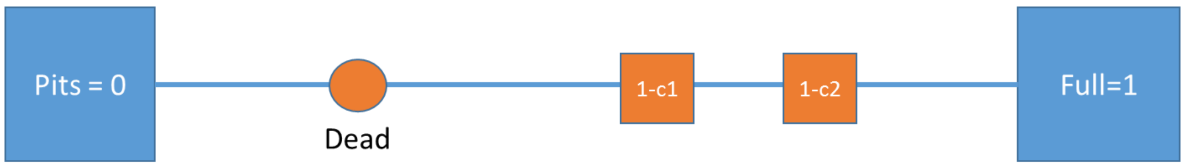
```
global_constant <- function(cornerStates){

#   Finds the global interaction constant.
#   Inputs: cornerStates -- list of corner states values
#   Outputs: coefficients of the degree len(cornerStates),
#   polynomial ($coefs) and roots (real and imaginary) of polynomial ($roots)
#   Notes: 0 will always be a root! Can handle ANY (real or imaginary) root!

  n <- length(cornerStates) # Number of attributes.
  poly.coefs <- rep(NA, n) # Vector to store the coefficients of the polynomial.
  for (i in 1:n) {
    # The coefficient on the degree i term is the sum over the products of
    # n choose i of the coefficients. (Save for the degree 1 term; below.)
    poly.coefs[i] <- sum(apply(combn(cornerStates, i), 2, prod))

  }
  poly.coefs[1] <- poly.coefs[1] - 1 # The degree 1 term needs a 1 subtracted from it.
  roots <- polyroot(poly.coefs) # Find the roots.
  result <- list(coefs = poly.coefs, roots = roots)
  return(result)
}
```

The sum of the c_is determines where to search for the root (Keeney & Raiffa, 1976; Torrance et al., 1996). If the sum is above 1, then $-1 < c < 0$. If the sum is below 1, then $c > 0$. If the sum is equal to 1, then $c = 0$, and the linear additive model holds.

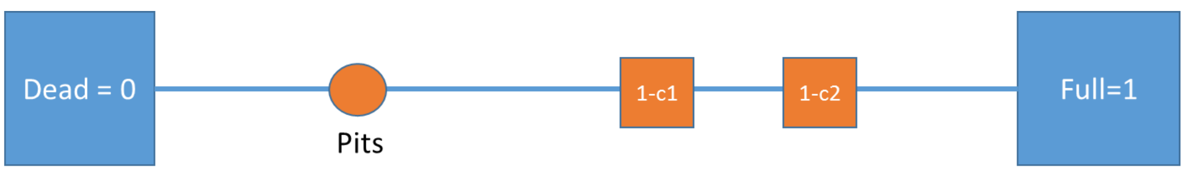
[2] The derivation of this scaling constant is explained in Step 9 of the Multi-attribute Function Estimation section, below.

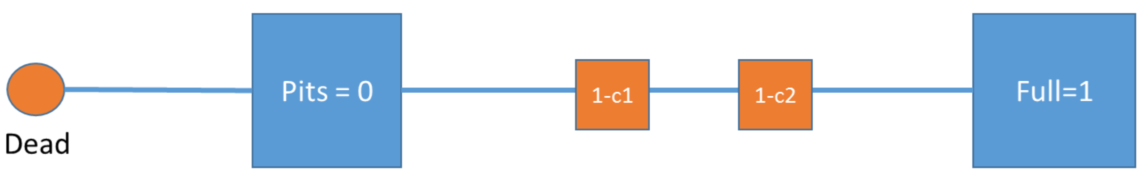
Multi-attribute Function Estimation

To build the multi-attribute function, it must first be constructed so that the utility of All-worst health state is 0. The disutility corner state values—the c_i in equation (1) —must be on this scale.[3] Participants could assign the bottom of the scale as All-worst health state ("Pits" in the figures) or Dead. If All-worst health state is the worst health state, rescaling is not needed (n=630):



If Dead is the worst health state, the disutility corner state values must be rescaled (n=354):





If we try to do this rescaling with individual respondents, many difficult situations arise. What do we do if All-worst health state is valued above a corner state? What if All-worst health state is valued the same as full health?

Rather than set up a complex and arbitrary set of exclusion and trimming rules, we decided to transform the mean scores within the "Dead is worst" group rather than transform individuals. Accordingly, we find the average value of All-worst health state in this group, the average utility values for the disutility

[3] As *dis*utilities, the c_i's are in fact on the scale where All-worst health state = 1 and Full Health = 0.

corner states for this group, and then use the former to transform the latter to the All-worst health state = 0 scale via an affine transformation. Below, we elaborate on this procedure, and describe the steps required to calculate the full multi-attribute function.

As noted elsewhere, for every mean value calculation, the 10% trimmed mean was used.

Step 1: Find the average utility values of the disutility corner states, for those who answered on the Dead = 0 scale (n=354).

Step 2: Transform the utility values from Step 1 to the All-worst health state = 0 scale.
Step 2 ensures that the responses of those who answered on the Dead = 0 scale (n=354) are commensurable with those who answered on the All-worst health state = 0 scale (n=630), the latter to be calculated in the next step. This step involves the following tasks:

a) Find the average utility value of All-worst health state among those who answered on the Dead = 0 scale.
b) Use the value from (a) to determine the affine transformation that keeps Full Health at 1 but moves All-worst health state to 0.
c) Apply the affine transformation from b) to the utilities from Step 1.

As described earlier, standard gamble responses that estimated a utility below 0 or above 1 were rounded up to 0 or down to 1, respectively. Step 2a) implemented a stricter constraint, rounding any utility above 0.5 down to 0.5. This was done to ensure that, upon transformation to the All-worst health state = 0 scale, the amount of utility space below 0 was equal to the amount of utility space above 0.

Step 2b) involves solving the following two equations:

$$\alpha + \beta \cdot u(pits) = 0$$
$$\alpha + \beta = 0$$

These equations move the utility of All-worst health state to 0, while maintaining the utility of Full Health at 1. Step 2c) involves plugging in the average disutility corner state utility values into the above equations, to get new utility values on the All-worst health state = 0 scale.

Step 3: Find the average utility values of the disutility corner states, for those who answered on the All-worst health state = 0.
Note that, in contrast to Step 2, Step 3 is straightforward, because these participants already provided their responses on the All-worst health state = 0 scale. No transformation is necessary.

Step 4: Compute the weighted average utility values of the disutility corner states.
To get the average utility values of the disutility corner states for the entire sample, average the values from Step 2c and Step 3, weighting the contribution of each sample by its size. The result is 7 disutility

corner state utility values, on the All-worst health state = 0 scale, that (collectively) use the responses of the whole sample.

Step 5: Compute 1 minus the values from Step 4, to get disutilities.
The values computed so far have been *utility* values. Subtracting each of the 7 from 1 gets us *disutility* values, on the scale where All-worst health state = 1 and Full Health = 0. Recall that the states in question are only true *corner* states in *disutility* space, and so the multi-attribute function must be built in terms of disutilities before being transformed back to a utility scale.

The disutility values are displayed in the table below (Table 13).

Table 13: Disutility values for the disutility corner states.

Domain	Disutility
Cognition	0.635
Depression	0.666
Fatigue	0.638
Pain	0.653
Physical Function	0.688
Sleep	0.562
Social Roles	0.611

Note, also, that the sum of the disutility corner state disutility values is 4.453. Thus, the linear additive model is rejected in favor of the multiplicative model. The more flexible multi-linear model might fit better, but the data requirements to estimate a multi-linear model were beyond the capability of this project.

Step 6: Using the values from Step 5, solve for the global interaction constant.

Refer to the equation and sample code in the previous section for how to solve the associated equation. The global interaction constant is -0.9992.

Step 7: Repeat Steps 2c-4 to find the average utility of Dead on the All-worst health state = 0 scale.

Ultimately, we need to have a utility score on the Dead = 0/Full Health = 1 scale to be able to combine morbidity and mortality. To estimate such a score, we need to perform a transformation akin to Step 2c) on the utilities produced from the multi-attribute function. To get the required constants for the transformation, we need to compute the utility of Dead on the All-worst health state = 0 scale. First, we find the average utility value of Dead for those who answered on the Dead = 0 scale. (This is trivial, as by definition, the value must be 0!) Then, transform it to the All-worst health state = 0 scale, using the transformation from Step 2c). Find the average utility value for those who answered on the All-worst health state = 0 scale, and then combine the two, weighting the contribution of each by its size.

The average utility of Dead on the All-worst health state = 0/Full Health = 1 scale is 0.0214.

Step 8: Compute 1 minus the value from Step 7, to convert it to a disutility.

As a disutility, the value from Step 7 becomes 0.979.

Step 9: Use the value from Step 8 to determine the scaling constant necessary to transform back to the Dead = 1/Full Health = 0 disutility scale.

This step involves computations akin to Step 2b). We solve the following equations:

$$\gamma + \kappa \cdot \bar{u}(dead) = 1$$
$$\gamma + \kappa \cdot 0 = 0$$

so that $\gamma = 0$ and $\kappa = \frac{1}{\bar{u}(dead)} = 1.02$.

Step 10: Combine Steps 5, 6, 9, and the single-attribute disutility functions, to get the multi-attribute scoring function.

We now have all the values to fill in equation (1):

- The single-attribute disutility functions come from the previous section.
- The disutility corner state values for c_i, on the All-worst health state = 1/Full Health = 0 disutility scale, come from Step 5.
- The global interaction constant c comes from Step 6.
- $\frac{1}{\bar{u}(dead)}$ comes from Step 9.

PROPr Scoring Function

The result of the above steps is the PROPr scoring function. The function takes as input a seven-element vector of theta scores:

$$\Theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7) = (\theta_{cognition}, \theta_{depression}, \theta_{fatigue}, \theta_{pain}, \theta_{physical}, \theta_{sleep}, \theta_{social})$$

Each of the theta scores are converted to a disutility estimate using the single attribute functions (see tables x-y):

$$(\theta_{cognition}, \theta_{depression}, \theta_{fatigue}, \theta_{pain}, \theta_{physical}, \theta_{sleep}, \theta_{social}) \xrightarrow[isotonic\ regression\ results]{} (CognitionDisutility, DepressionDisutility, FatigueDisutility, PainDisutility, PhysicalDisutility, SleepDisutility, SocialDisutility)$$

These disutility estimates are combined with the global interaction constant, disutility corner states, and $\frac{1}{\bar{u}(dead)}$ to estimate a single PROPr score:

$$PROPr = 1 - 1.0219(\frac{1}{-0.99918} \times (1 + -0.99918 \times 0.63504 \times CognitiveDisutility) \times (1 + -0.99918 \times 0.66616 \times DepressionDisutility) \times (1 + -0.99918 \times 0.63861 \times FatigueDisutility) \times (1 + -0.99918 \times 0.65296 \times PainDisutility) \times (1 + -0.99918 \times 0.68835 \times PhysicalDisutility) \times (1 + -0.99918 \times 0.56296 \times SleepDisutility) \times (1 + -0.99918 \times 0.61126 \times SocialDisutility))$$

and outputs a utility on the Dead = 0/Full Health = 1 scale, where 1 is the highest utility, and utilities less than 0—for states valued as worse than Dead (minimum score of-0.022)—are possible.

PROPr Scoring Function Example

A person has completed items from PROMIS item banks that result in the following domain scores:

Domain	T-score	Theta score (=(t-score-50)/10)
Cognition	56.1	0.61
Depression	41	-0.9
Fatigue	33.7	-1.63
Pain Interference	41.6	-0.84
Physical Function	57	0.7
Sleep Disturbance	41.2	-0.88
Ability to Participate in Social Roles and Activites	51.8	0.18

Table 14: Worked example theta scores

Using the isotonic regression results, these theta scores can be converted into disutility estimates (see Table 6 to Table 12):

Domain	T-score	Theta score	Disutility Estimate
Cognition	56.1	0.61	0.1062
Depression	41	-0.9	0.0304
Fatigue	33.7	-1.63	0.0814
Pain Interference	41.6	-0.84	0
Physical Function	57	0.7	0.0615
Sleep Disturbance	41.2	-0.88	0.1246
Ability to Participate in Social Roles and Activites	51.8	0.18	0.1833

Table 15: Worked example disutility estimates

Visually, this conversion is illustrated in Figure 12. The theta score (blue line) is associated with a segment of the isotonic regression results (black line). Using the isotonic regression results tables, this can be used to estimate a disutility score (orange line).

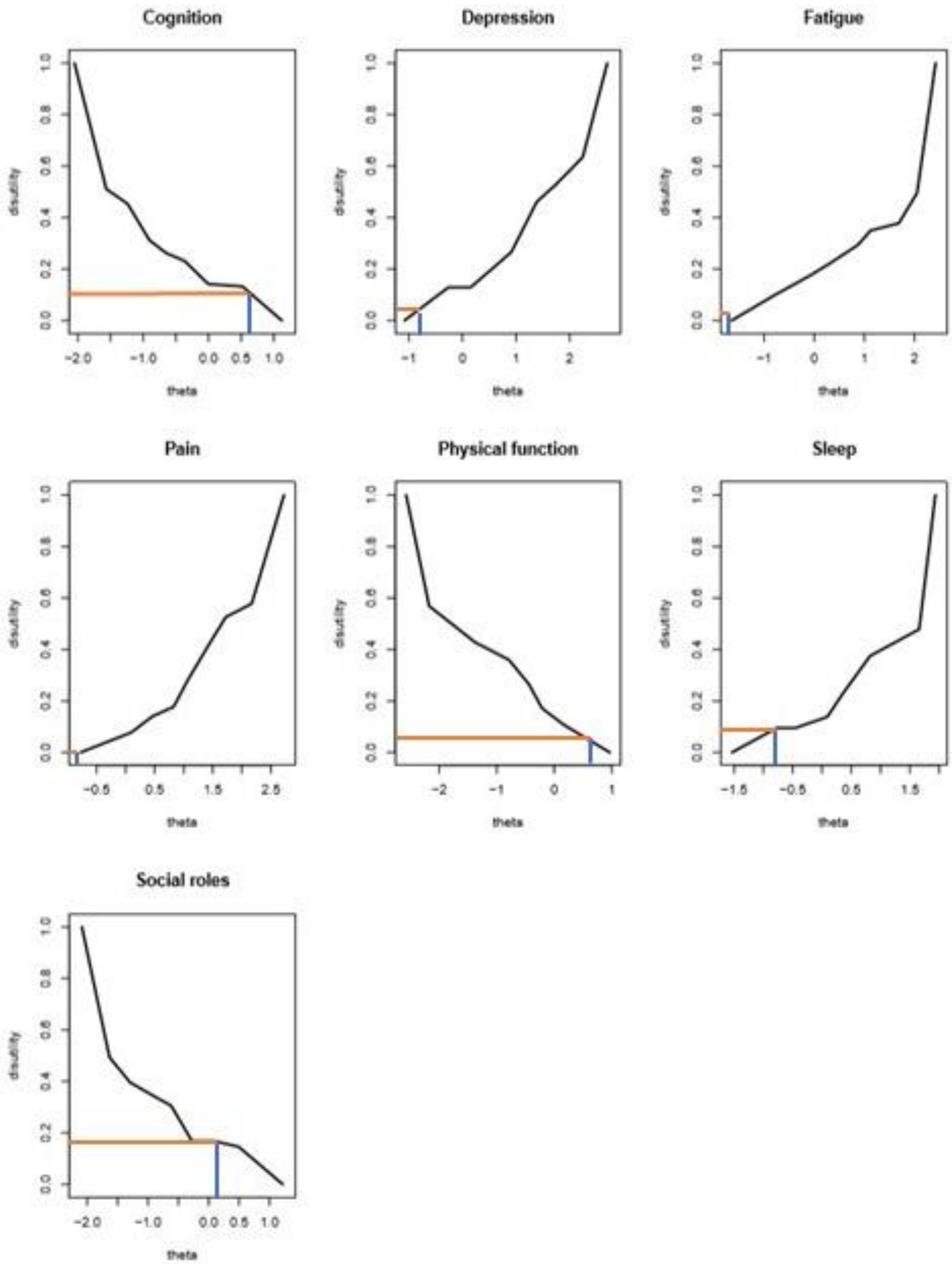


Figure 12: Visualization of the theta scores to disutility scores conversion

The single attribute disutility scores can be converted to single attribute utility scores (because utility = 1 – disutility):

Domain	T-score	Theta score	Disutility Estimate	Utility Estimate
Cognition	56.1	0.61	0.1062	0.8937
Depression	41	-0.9	0.0304	0.9695
Fatigue	33.7	-1.63	0.0814	0.9185
Pain Interference	41.6	-0.84	0	1
Physical Function	57	0.7	0.0615	0.9384
Sleep Disturbance	41.2	-0.88	0.1246	0.8753
Ability to Participate in Social Roles and Activites	51.8	0.18	0.1833	0.8166

Table 16: Worked example utility estimates

The single attribute disutility scores can also be combined with the global interaction constant, disutility corner states, and $\frac{1}{\bar{u}(dead)}$ to estimate a single PROPr score:

$$
\begin{aligned}
PROPr = 1 - 1.0219 \Bigg(& \frac{1}{-0.99918} \times (1 + -0.99918 \times 0.63504 \times 0.1062) \\
& \times (1 + -0.99918 \times 0.66616 \times 0.0304) \times (1 + -0.99918 \times 0.63861 \times 0.0814) \\
& \times (1 + -0.99918 \times 0.65296 \times 0) \times (1 + -0.99918 \times 0.68835 \times 0.0615) \\
& \times (1 + -0.99918 \times 0.56296 \times 0.1246) \times (1 + -0.99918 \times 0.61126 \times 0.1833) \Bigg) \\
= {} & 0.757
\end{aligned}
$$

We have provided standardized code in the Appendixes for R and SAS to do these calculations.

Sensitivity Analysis

There are many design choices made in the creation of the PROPr scoring function. To test the importance of these choices, we recalculated the preference-base scoring functions, using the following alternative cases:

i. Changing the sample to one with no minimum time threshold, applying 10% trimming during calculations.
ii. Changing the sample to one with the minimum time threshold, without 10% trimming.
iii. Changing the sample to one with the minimum time threshold, removing utility responses that were estimated below 0 or above 1.
iv. Changing the sample to a "selective" sample, described below.
v. Maintaining our base case sample, but altering the disutility corner state calculations, as described below.

Case (i) refers to the sample of *n* = 1164, applying 10% trimming at every calculation (as was done in the base case). Case (ii) maintains the base case sample of *n* = 983, but does not perform 10% trimming. Case (iii) maintains the base case sample of *n* = 983, but removes individual utility responses that were estimated below 0 or above 1. Recall that, previously, these were rounded to 0 and 1, respectively. Case (iv) refers to a sample defined by the exclusion of anyone who met any of the following criteria: completed the survey in under 15 minutes; violated monotonicity by more than 10% of the scale, more than twice; used less than 10% of the scale for all of their valuations; or, rated their understanding of the survey as less than 2 on a scale of 1 = "Not at all" to 6 = "Very much." These criteria are based on the variety of exclusions used in the literature (Engel, Bansback, Bryan, Doyle-Waters, & Whitehurst, 2015). Case (v) refers to maintaining the base case sample of *n* = 983, but altering the procedure used to calculate the disutility corner state values. For a given disutility corner state, rather than including anyone who valued that state in determining its (dis)utility, we included only those who did *not* also value the domain to which the disutility corner state belongs. The rationale for case (v) was to see if only including the contribution of those who had valued some other domain beforehand—and thus might be contrasting that domain with the domain of the disutility corner state under consideration—would have an aggregate effect on the scoring function.

We computed the PROPr scoring function using the alternative cases described above. We then evaluated these functions, as well as the original, on the sample of 983 participants described earlier. (Recall that participants were asked enough questions from PROMIS for their HRQL to be evaluated using the scoring function their responses—on the elicitation questions—were helping to create.) We found that the utilities of the sample produced using the original function and those produced using any of the alternatives were correlated above 0.98, and that, for cases (i)-(iv), the rankings of the disutility corner states—giving the relative importance of the health domains—are the same until the 4th or 5th position. For case (v), the ranking was pain, physical functioning, cognition, depression, social roles, fatigue, sleep, whereas the original ranking (see Table 13) is physical functioning, depression, pain, fatigue, cognition, social roles, sleep.

Quantification of Uncertainty

PROPr, through PROMIS, offers an interesting opportunity for quantifying the uncertainty around utility scores. The theta values corresponding to the health-state descriptions in the PROPr survey (Table 17) are, in fact, point estimates, with known precision calculated through IRT. These are displayed in the table below.

Table 17: Theta estimates and standard deviation of each estimate by domain

Cognition		Depression		Fatigue		Pain Interference		Physical Function		Sleep Disturbance		Social Roles	
Theta	SD	Theta	SD	Theta	SD	Theta	SD	Theta	SD	Theta	SD	Theta	SD
1.124	0.689	-1.082	0.617	-1.648	0.611	-0.773	0.67	0.966	0.666	-1.535	0.687	1.221	0.576
0.52	0.591	-0.264	0.426	-0.818	0.506	0.1	0.404	0.16	0.474	-0.775	0.615	0.494	0.372
-0.002	0.524	0.151	0.385	-0.094	0.487	0.462	0.413	-0.211	0.451	-0.459	0.592	0.083	0.341
-0.367	0.55	0.596	0.384	0.303	0.447	0.827	0.331	-0.443	0.443	0.093	0.568	-0.276	0.337
-0.649	0.521	0.913	0.397	0.87	0.436	1.072	0.349	-0.787	0.49	0.335	0.569	-0.618	0.347
-0.902	0.534	1.388	0.398	1.124	0.441	1.407	0.345	-1.377	0.417	0.82	0.584	-0.955	0.328
-1.239	0.541	1.742	0.406	1.688	0.48	1.724	0.368	-1.784	0.465	1.659	0.659	-1.293	0.378
-1.565	0.572	2.245	0.43	2.053	0.508	2.169	0.404	-2.174	0.526	1.934	0.699	-1.634	0.363
-2.052	0.646	2.703	0.479	2.423	0.61	2.725	0.492	-2.575	0.622			-2.088	0.493

Validity of the Preference-Based Scoring System

The preference-based scoring system's validity has been established in three ways:

1. Known-groups construct validity: Respondents with a chronic health condition report lower PROPr scores than those without the health condition.
2. Convergent validity: PROPr scores are correlated with scores from previously developed measures (the EQ-5D-5L, the HUI2, and the HUI3).
3. Known-groups construct validity: PROPr scores discriminate between groups (such as age groups and gender) in the same way as previously developed measures.

This validity has been established using both the data described above (the PROPr survey) and the PROMIS Rescaling Survey. The PROPr survey administered the PROMIS-29, Cognition 4-item short form, EQ-5D-5L, and self-administered HUI2 and HUI3. The Rescaling survey administered 10 PROMIS domains (with 8 to 13 items per domain) and the self-administered HUI2 and HUI3.

The PROPr survey included the EQ-5D-5L for which lacked US scoring. We used the EQ-5D-5L to EQ-5D-3L US scoring look-up table to score the EQ-5D-5L for this technical report. Both the PROPr survey and the Rescaling survey included the self-administered HUI which was scored for Mark 2 and Mark 3.

PROMIS Domain Information in Each Survey

The PROPr survey only included PROMIS short forms (4 items per domain). We assigned theta scores using the Assessment Center Scoring Service. Calibration samples for scoring were the Cancer Supplement Calibration (Cognition), PROMIS Wave 1 (Depression, Fatigue, Pain), PROMIS Wave 1 with Extension (Physical Function), PROMIS Sleep Wave 1 (Sleep), and Social Supplement (Social Roles).

Because the short forms and look-up tables were used for scoring the PROMIS domains, the range of possible PROMIS theta scores is limited.

With the 4-item short forms, the best possible theta scores are not as healthy as the theta scores necessary to have the best PROPr score. Table 18 contains the theta scores necessary to get a PROPr score of 1.0 and the maximum achievable theta score using the 4-item short forms. The rightmost column shows that the best health state in the PROPr scoring system cannot be reached using the 4-item short form for 3 of the 7 PROPr domains. The maximum PROPr score using the PROMIS 4-item short forms is 0.955; this maximum score was observed in the PROPr Survey.

	PROPr Full Health Theta	**PROMIS Short Form Full Health Theta**	**Short form reaches PROPr best?**
Cognition	1.124	1.38	Yes
Depression	-1.082	-0.90	No
Fatigue	-1.648	-1.63	No
Pain	-0.773	-0.84	Yes
Physical Function	0.966	0.69	No
Sleep	-1.535	-1.80	Yes
Social Roles	1.221	1.42	Yes

Table 18: Highest health theta scores in PROPr and the 4-item standardized short forms

There is a similar limitation when comparing the worst health states reportable in the PROMIS 4-item short forms and the PROPr scoring system. Table 19 contains the theta scores necessary to get a PROPr score of -0.022 and the minimum achievable theta score using the 4-item short forms. The rightmost column shows that the worst health state in the PROPr scoring system cannot be reached using the 4-item short form for 2 of the PROPr domains. The minimum PROPr score using the PROMIS 4-item short forms is -0.019; this minimum score was observed in the PROPr Survey.

	PROPr Worst Health Theta	**PROMIS Short Form Worst Health Theta**	**Short form reaches PROPr worst?**
Cognition	-2.052	-1.99	No
Depression	2.703	2.94	Yes
Fatigue	2.423	2.58	Yes
Pain	2.725	2.56	No
Physical Function	-2.575	-2.71	Yes
Sleep	1.934	2.33	Yes
Social Roles	-2.088	-2.25	Yes

Table 19: Lowest health theta scores in PROPr and the 4-item standardized short forms

The Rescaling survey included 8 to 13 items per domain. We assigned theta scores using the Assessment Center Scoring Service. Calibration samples for scoring were the Cancer Supplement Calibration (Cognition), PROMIS Wave 1 (Depression, Fatigue, Pain), PROMIS Wave 1 with Extension (Physical Function), PROMIS Sleep Wave 1 (Sleep), and Social Supplement (Social Roles). The items covered the full range of possible PROPr scores and the full range of possible PROPr scores was observed.

Known-groups construct validity: Respondents with a chronic health condition report lower PROPr scores than those without the health condition

Figure 13 illustrates age- and sex-adjusted condition impact estimates for 11 chronic health conditions from the PROPr survey. Participants were asked about these health conditions using the standardized language from NHIS (e.g., "Have you EVER been told by a doctor or other health professional that you had coronary heart disease?").

Condition impact estimates were created using ordinary least squares regression:

$$PROPr_i = \beta_0 + \beta_1(age_i - 50) + \beta_2(female_i) + \beta_3(condition_i) + \varepsilon_i$$

A separate analysis was done for each condition; β_3 is the condition impact estimate.

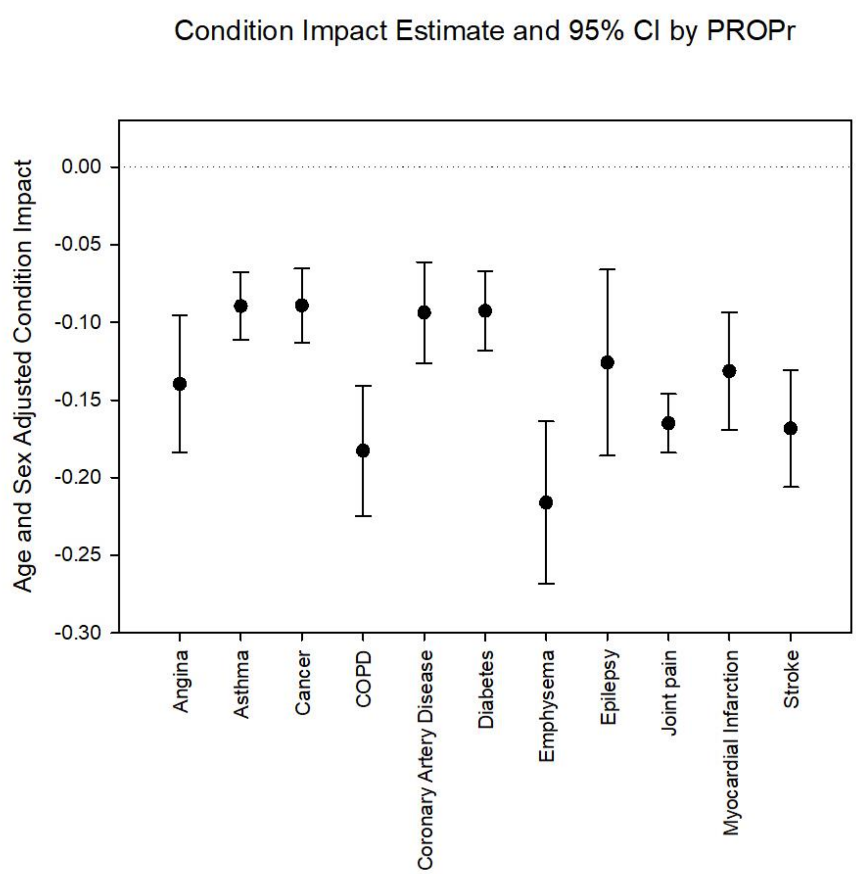


Figure 13: PROPr Survey, condition impact estimates for PROPr

All chronic health conditions have a statistically significant impact on PROPr scores. Note that this survey did not ask about disease severity.

Figure 14 illustrates age- and sex-adjusted condition impact estimates for 21 chronic health conditions from the Rescaling survey. Participants were asked about these health conditions using language similar to NHIS (e.g., “Have you ever been told by a doctor or a health professional that you had any of the following? . . .”).

Condition impact estimates were created using ordinary least squares regression:

$$PROPr_i = \beta_0 + \beta_1(age_i - 50) + \beta_2(female_i) + \beta_3(condition_i) + \varepsilon_i$$

A separate analysis was done for each condition; $\beta_3$is the condition impact estimate.

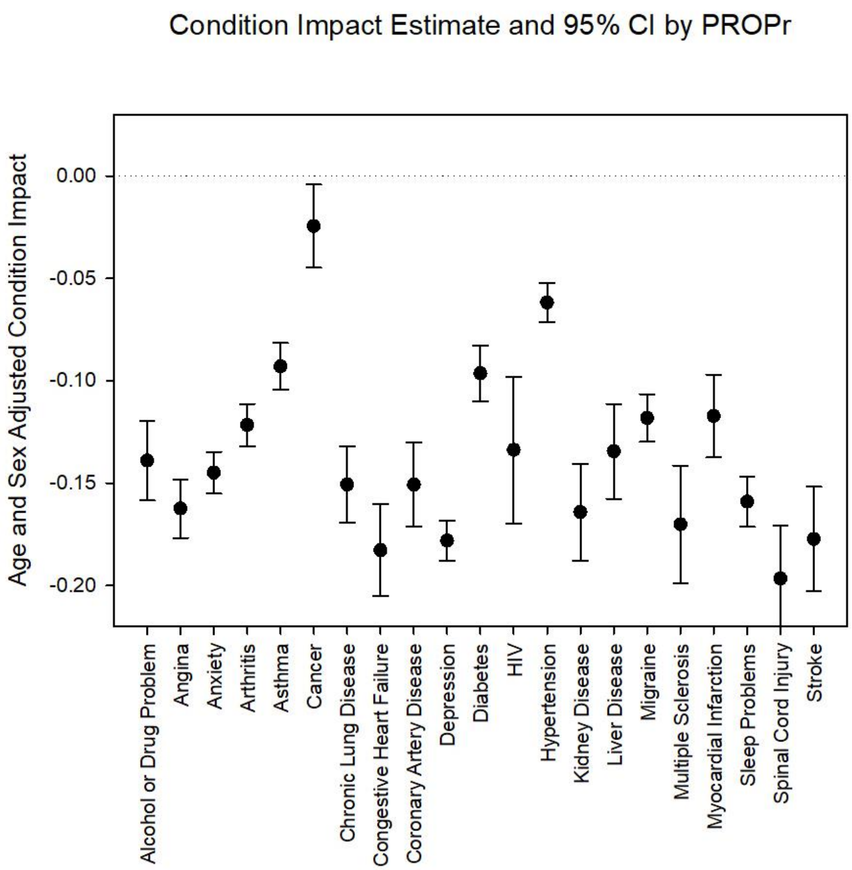


Figure 14: Rescaling survey, condition impact estimates for PROPr

All chronic health conditions have a statistically significant impact on PROPr scores. Note that this survey did not ask about disease severity.

Convergent validity: PROPr scores are correlated with scores from previously developed measures (the EQ-5D and the HUI)

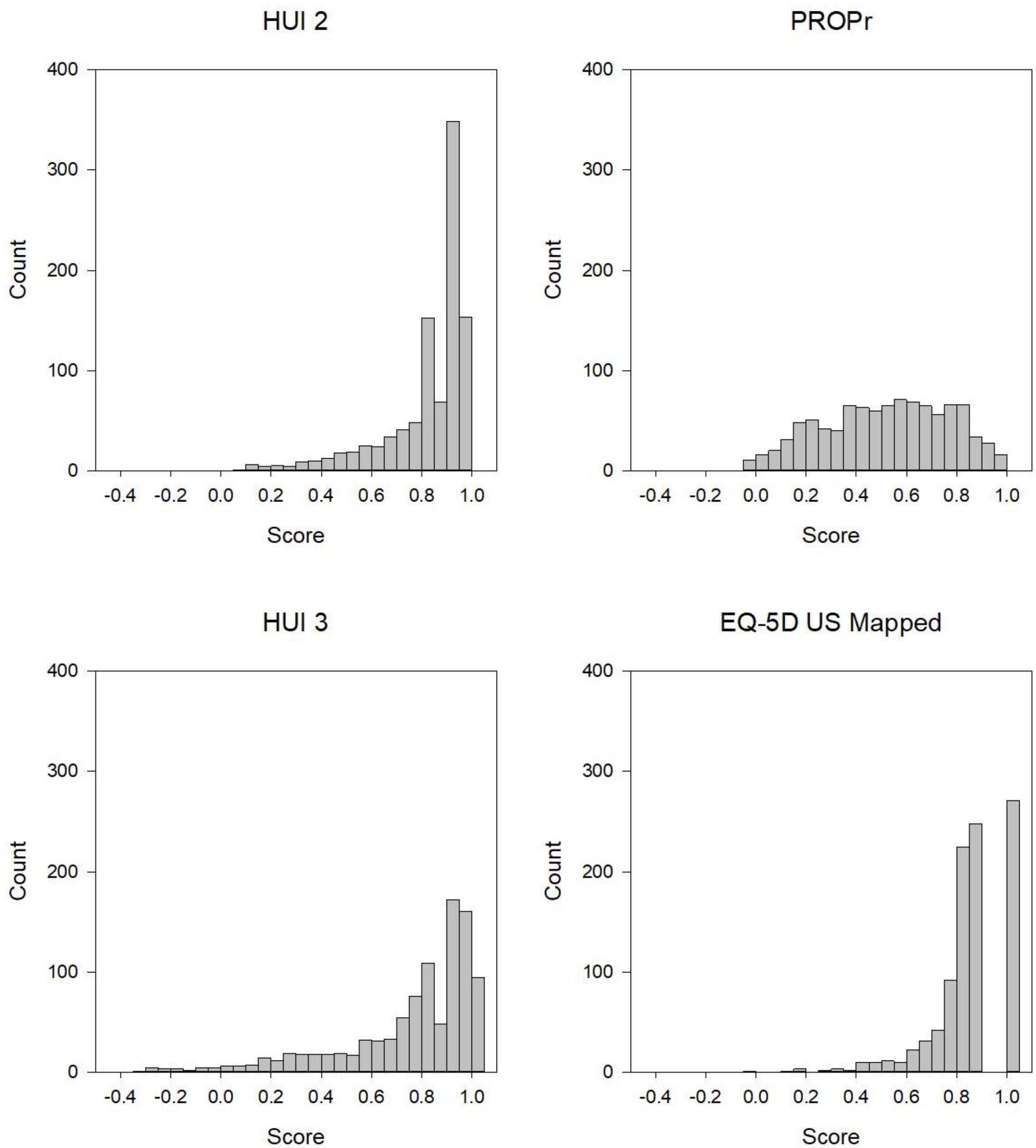


Figure 15: PROPr survey, histograms of all summary scores

Figure 15 has histograms of all the summary scores collected in the PROPr survey: PROPr, EQ-5D-5L mapped to US scoring, HUI Mark 2, and HUI Mark 3. The EQ-5D-5L, HUI Mark 2, and HUI Mark 3 all have a ceiling effect in this general population sample. PROPr, which does not appear to have a ceiling or floor effect in this sample, has a lower mean score than any of the other systems.

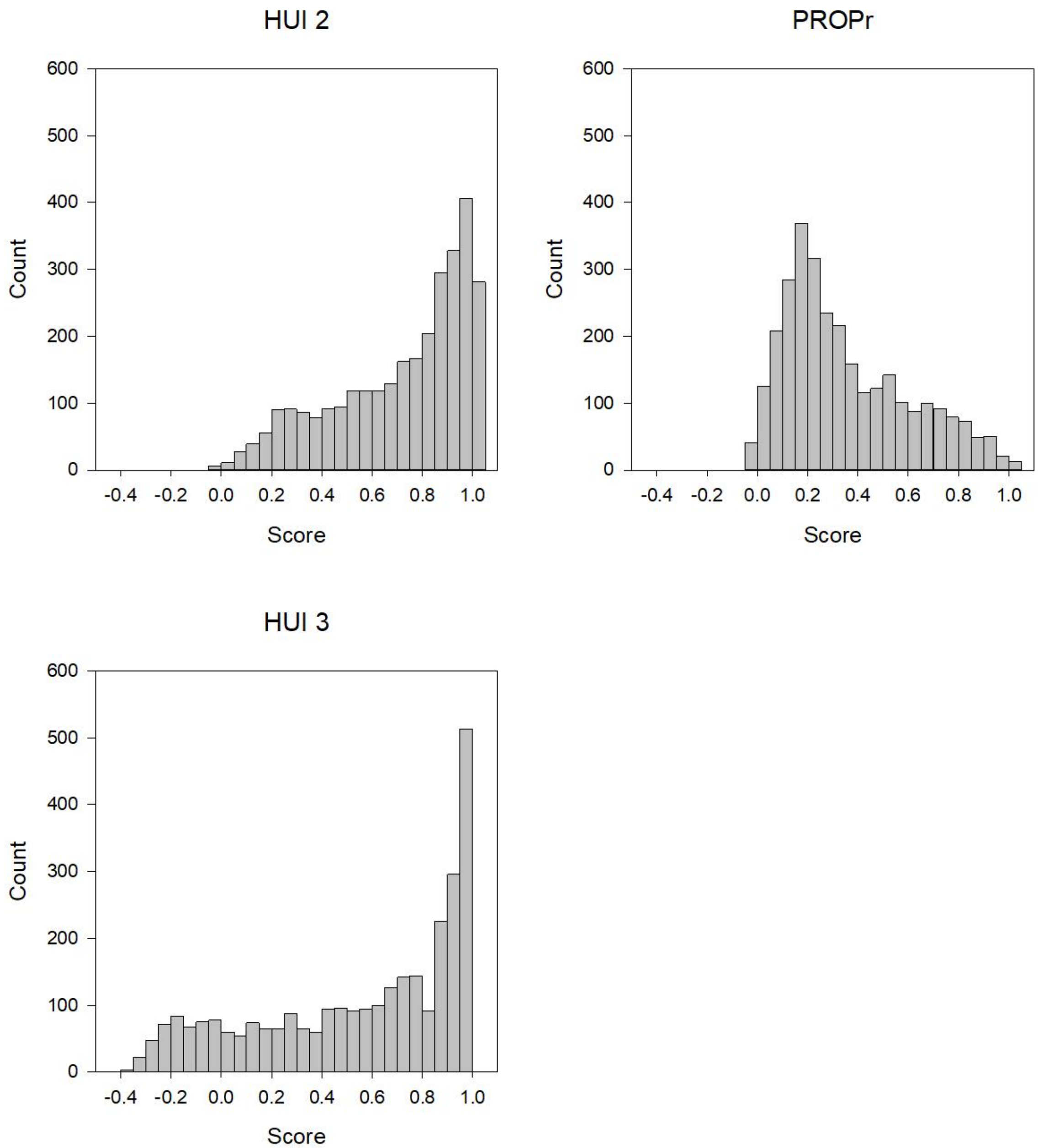


Figure 16: Rescaling survey, histograms of all summary scores

Figure 16 has histograms of all the summary scores collected in the Rescaling survey: PROPr, HUI Mark 2, and HUI Mark 3. The HUI Mark 2 and HUI Mark 3 have a ceiling effect in this general population sample. PROPr does not appear to have a ceiling effect in this sample; PROPr has a lower mean score than the other systems.

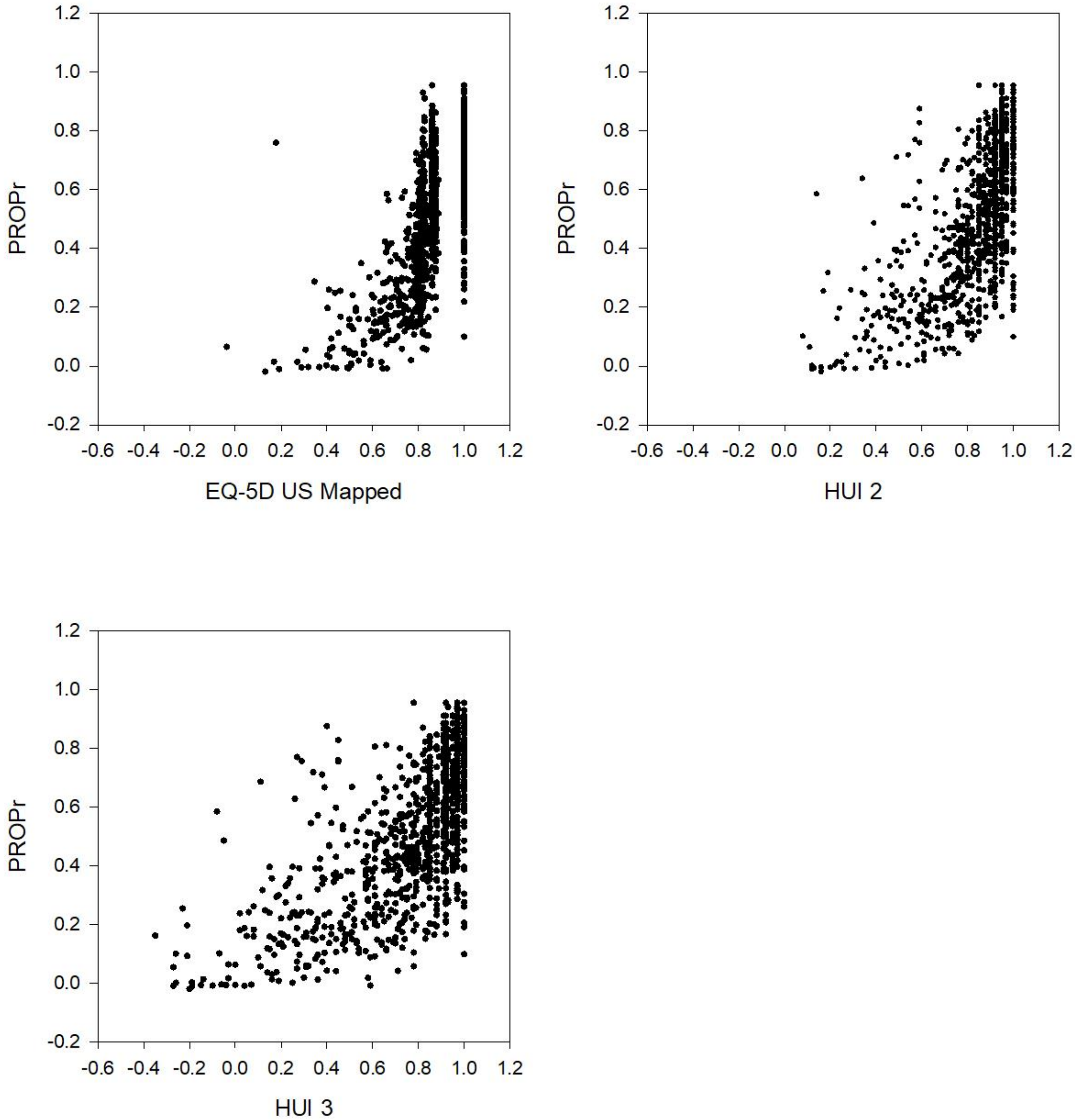


Figure 17: PROPr survey, scatterplots of all summary scores

Figure 17 has scatterplots of participant scores for PROPr and the other summary measures in the PROPr survey. Visually, the scores appear correlated with a ceiling effect in the EQ-5D-5L, HUI Mark 2, and HUI Mark 3.

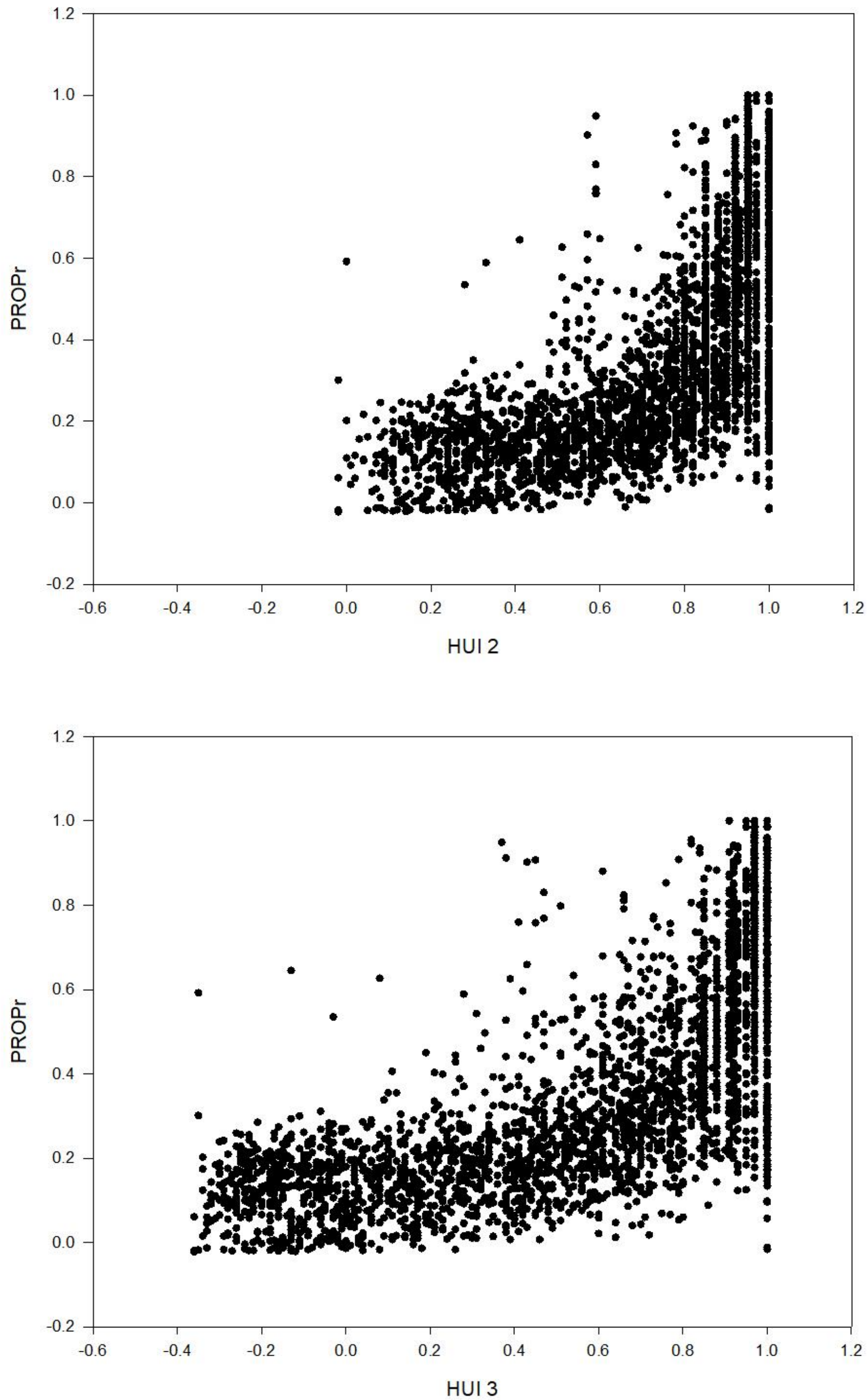


Figure 18: Rescaling survey, scatterplots of all summary scores

Figure 18 has scatterplots of participant scores for PROPr and the other summary measures in the Rescaling survey. Visually, the scores appear correlated with a ceiling effect in the HUI Mark 2 and HUI Mark 3.

Table 20 and Table 21 report the Pearson correlations between summary scores in the PROPr survey and Rescaling survey, respectively. Note that the correlations between HUI Mark 2 and HUI Mark 3 are artificially inflated because some of the same questions were used to create both scores.

	HUI2	**HUI3**	**PROPr**
EQ-5D	.720	0.702	0.703
HUI2		0.905	0.661
HUI3			0.673

Table 20: PROPr survey, correlations of summary scores

	HUI3	**PROPr**
HUI2	0.934	0.666
HUI3		0.696

Table 21: Rescaling survey, correlations of summary scores

In these samples, the correlation between PROPr and the other summary measures ranges from 0.661 to 0.703. A prior study which has co-administered preference-based summary measures found correlations from 0.60 to 0.71 [Fryback 2007].

Known-groups construct validity: PROPr scores discriminate between groups in the same way as previously developed measures

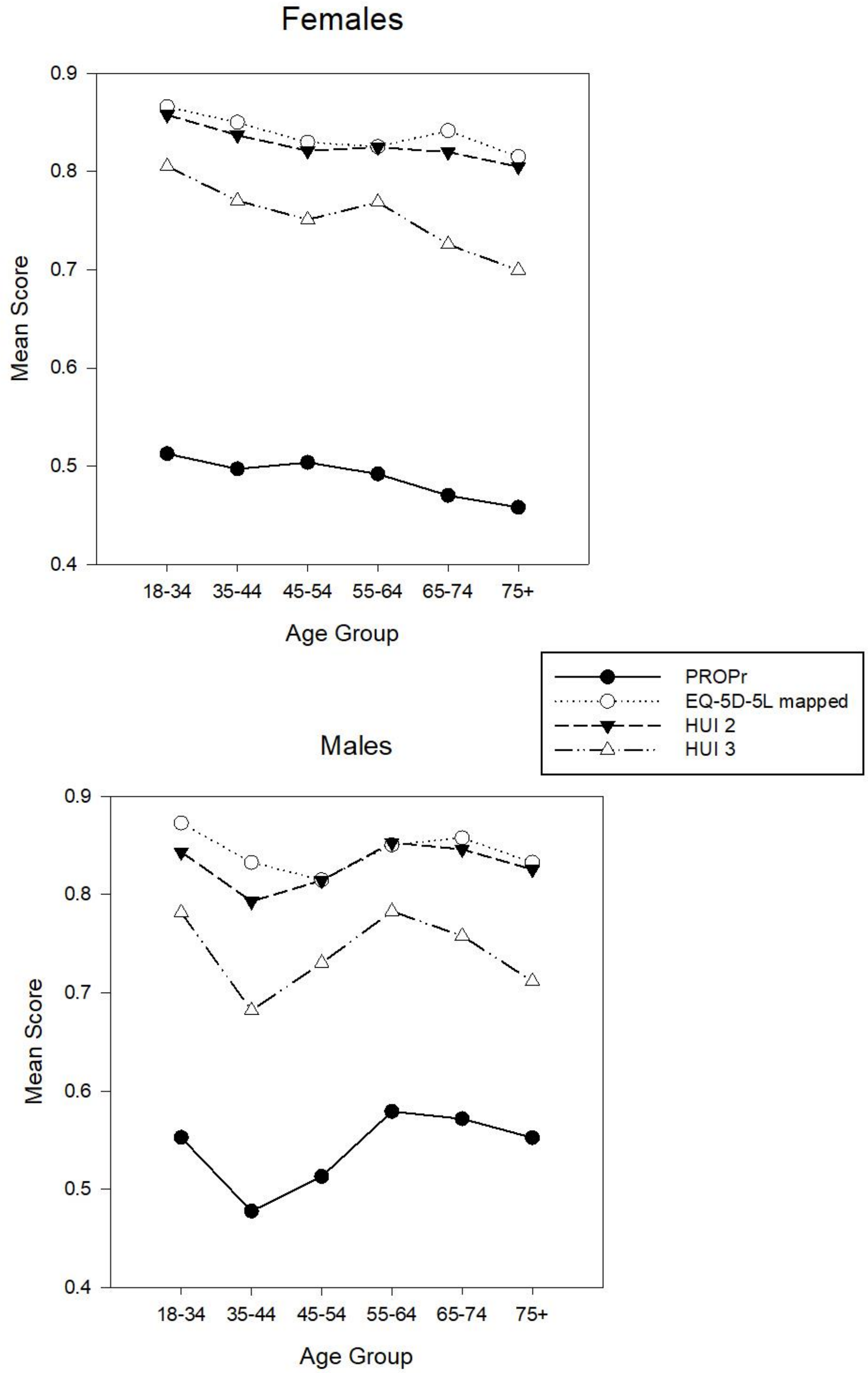


Figure 19: PROPr survey, age- and sex-stratified mean scores for all summary scores

Figure 19 illustrates age- and sex-stratified mean scores for all summary scores collected in the PROPr survey. The scores mimic each other, with the PROPr scores having lower values than the other scores.

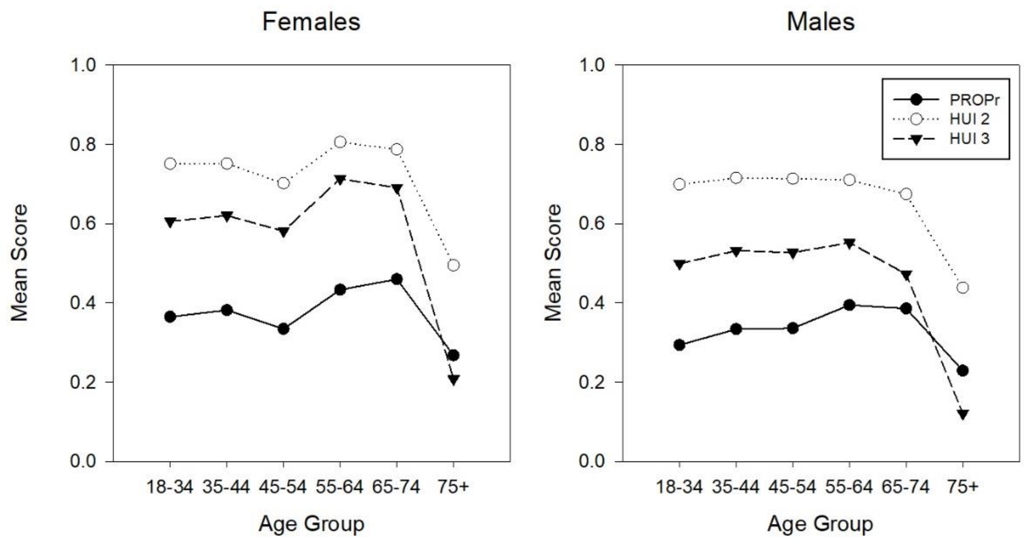


Figure 20: Rescaling survey, age- and sex-stratified mean scores for all summary scores

Figure 20 illustrates age- and sex-stratified mean scores for all summary scores collected in the Rescaling survey. The scores mimic each other, with the PROPr scores having lower values than the other scores except in the oldest male age strata.

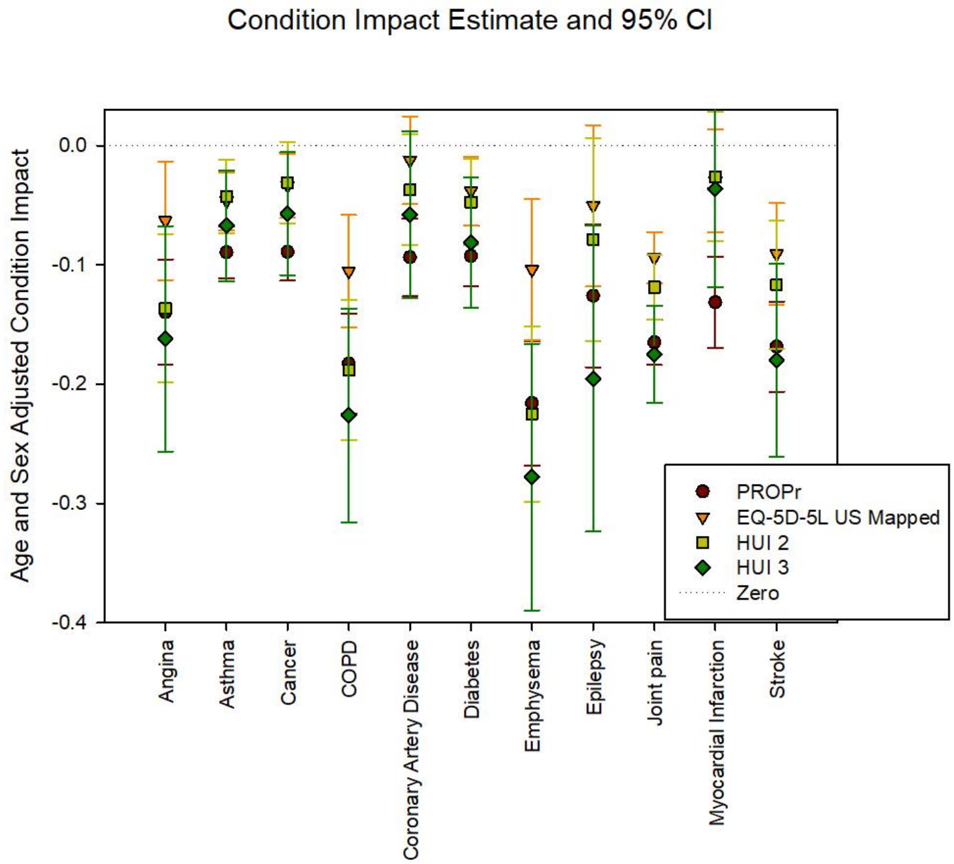


Figure 21: PROPr survey, condition impact estimates for all summary scores

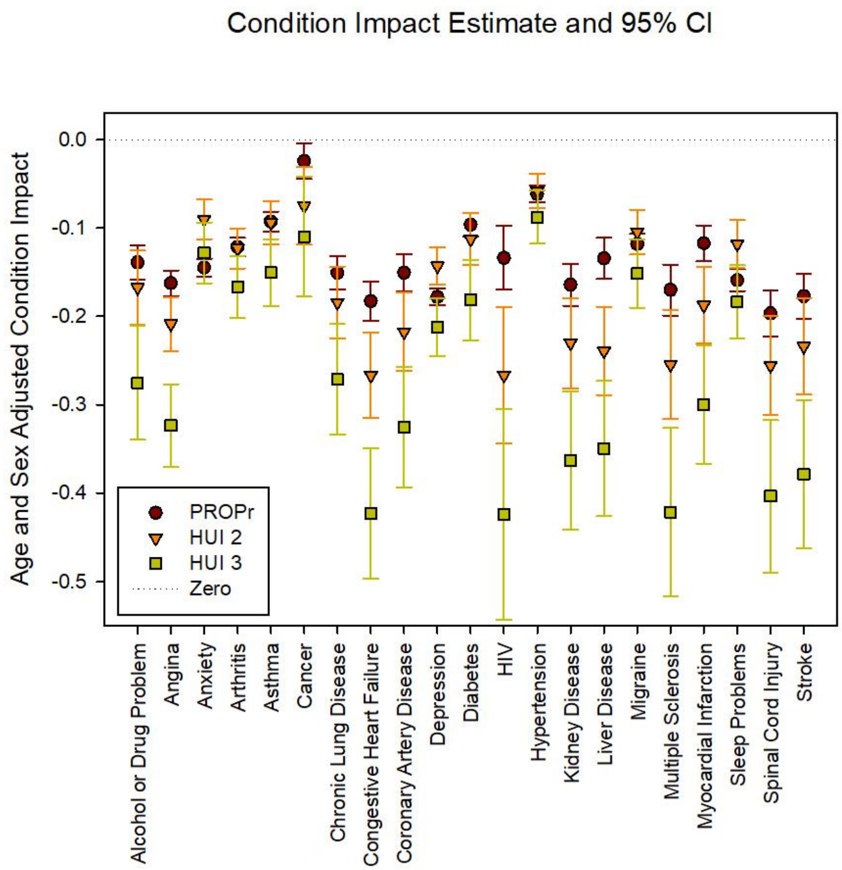


Figure 22: Rescaling survey, condition impact estimates for all summary scores

Figure 21 and Figure 22 illustrate age- and sex-adjusted chronic condition impact estimates for all summary measures in the PROPr survey and Rescaling survey, respectively.

In the PROPr survey, all 11 condition impact estimates were statistically significant by PROPr, 8 were statistically significant by the EQ-5D-5L, 7 were statistically significant by HUI Mark 2, and 9 were statistically significant by HUI Mark 3. In the Rescaling survey, all 21 condition impact estimates were statistically significant of all summary scores by all three scoring systems.

The order of conditions by impact estimate is similar across all summary scores. The HUI Mark 3 generally has the largest impact estimates**.** In the PROPr survey, condition impact estimates by PROPr are most similar to the HUI Mark 3 for conditions with less impact, and most similar to the HUI Mark 2 for conditions with more impact. In contrast, in the Rescaling survey, condition impact estimates by PROPr are most similar to the HUI Mark 2.

Discussion

This technical report has given an overview of the motivation, design, estimation, and evidence for the PROPr preference-based summary scoring system for health-related quality of life. PROPr was carefully constructed with input from a variety of measurement experts as well as community members. The final scoring algorithm was estimated using a large representative sample of the US noninstitutionalized population.

PROPr is the first score to link directly single-attribute utility functions to health domains as measured by Item Response Theory (IRT). As such, PROPr should gain many of the advantages of an IRT-based descriptive system, including flexible administration of items from the 7 item banks used to construct PROPr. A PROPr score can be estimated if there are score estimates from each of the 7 PROMIS domains; a PROMIS domain score can be estimated with as little as a single item, though such estimates have limited range and precision. Ideally, a user would select a set of items that would allow a respondent to get a domain score over the entire range of domain scores covered by PROPr. If a reduced range of domain scores is available to the respondent, the respondent will also have a reduced range of PROPr scores (see Table 18 and Table 19). The PROPr score can still be calculated, but will be artificially truncated.

PROPr scores are lower than EQ-5D-5L, HUI2, and HUI3 scores in the samples where they were co-administered. We expected PROPr scores to be lower than these legacy measures because the best health state described in PROPr is significantly better than the best health state described in legacy measures. For example, the best physical functioning in PROPr is “able to dress yourself, including tying shoelaces and buttoning up your clothes without any difficulty *and* able to run 100 yards (100 m) without any difficulty.” In contrast, the best physical functioning in the EQ-5D-5L is “I have no problems walking,” and in the HUI Mark 3, “I have full use of two hands and ten fingers *and* I am able to walk around the neighbourhood without difficulty, and without walking equipment.” The increase in descriptive space in PROPr “raises the bar” to reach a best-health score of 1.0. This will both decrease ceiling effects in the general population and substantially lower scores when PROPr is compared to legacy measures.

Even though PROPr has lower absolute scores than the EQ-5D-5, HUI2, or HUI3, the difference between scores for groups with and without different chronic medical conditions as measured by PROPr is between the difference in scores seen is most similar to the HUI Mark 2 (see Figure 21 and Figure 22). The estimated impact of the chronic conditions is both in the expected direction (the group with a condition has a lower mean than the group without that condition) and statistically significant. Future work will be needed to establish thresholds for a clinically meaningful difference in PROPr scores; most preference-based measures have clinically meaningful difference thresholds between 0.03 and 0.05, although a conservative estimate of 0.08 (half of a standard deviation in the PROPr dataset) could also be used. Future work is also needed to collect full PROPr scores from a nationally representative population, create crosswalks to legacy measures, and validate PROPr using longitudinal data collection.

The findings presented in this report provide a solid foundation for the use of PROPr. The single-attribute functions can be used as cardinal measures of utility in each domain. The multi-attribute function on the Dead = 0 and Full Health = 1.0 scale allows PROPr scores to be used for calculating aggregated indices of morbidity and mortality (such as quality-adjusted life years). In summary, the PROPr score is a comprehensive and efficient system for describing general health status and estimating a preference-based summary score across a wide variety of subjects and studies.

References

Craig, B.M., Reeve, B.B., Brown, P.M., Cella, D., Lipscomb, J., Pickard, S., Revicki, D.A. (2014). US valuation of health outcomes measured using the PROMIS-29. Value in Health, 17(8), 846-853.

Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., Denton, M. and Boyle, M., Multiattribute and single-attribute utility functions for the health utilities index mark 3 system, Medical care, Vol. 40, No. 2, pp. 113–128, 2002.

Fryback, D. G., Dunham, N. C., Palta, M., Hanmer, J., Buechner, J., Cherepanov, D., ... & Feeny, D. (2007). US norms for six generic health-related quality-of-life indexes from the National Health Measurement study. Medical care, 45(12), 1162.

Furlong, W., Feeny, D., Torrance, G., Goldsmith, C., DePauw, S., Zhu, Z., ... & Boyle, M. (1998). Multiplicative multi-attribute utility function for the Health Utilities Index Mark 3 (HUI3) system: a technical report (No. 1998-11). Centre for Health Economics and Policy Analysis (CHEPA), McMaster University, Hamilton, Canada.

Hanmer, J., Feeny, D., Fischhoff, B., Hays, R. D., Hess, R., Pilkonis, P. A., ... & Yu, L. (2015). The PROMIS of QALYs. Health and quality of life outcomes, 13(1), 1.

Hanmer, J., Cella, D., Feeny, D., Fischhoff, B., Hays, R. D., Hess, R., ... & Yu, L. (2017). Selection of key health domains from PROMIS® for a generic preference-based scoring system. *Quality of Life Research*, *26*(12), 3377-3385.

Hays, R.D., Revicki, D.A., Feeny, D., Fayers, P., Spritzer, K.L., Cella, D. Using linear equating to map PROMIS global health items and the PROMIS-29 V2.0-profile measure to the Health Utilities Index Mark 3. Pharmacoeconomics 2016 (ePub).

Herdman, M., Gudex, C., Lloyd, A., Janssen, M. F., Kind, P., Parkin, D., ... & Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of life research, 20(10), 1727-1736.

Keeney, R. L. and Raiffa, H., Decision with multiple objectives: Preferences and value tradeoffs, John Wiley & Sons, New York, NY, 1976.

Revicki, D.A., Kawata, A.K., Harnam, N., Chen, W.H., Hays, R.D., Cella, D. (2009). Predicting EuroQol (EQ-5D) scores from the patient-reported outcomes measurement information system (PROMIS) global items and domain item banks in a United States sample. Qual Life Res, 18(6), 783-791.

Torrance, G. W., Feeny, D. H., Furlong, W. J., Barr, R. D., Zhang, Y. and Wang, Q., Multiattribute utility function for a comprehensive health status classification system: Health Utilities Index Mark 2, Medical care, Vol. 34, No. 7, pp. 702–722, 1996.

Shalizi, C., Advanced Data Analysis from an Elementary Point of View, Cambridge University Press, Cambridge, UK, forthcoming, http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV.

Appendix 1: SAS Code

```
 /*  DESCRIPTION
   This is the multi-attribute utility function, using isotonic regression
with linear interpolation to model the single-attribute (dis)utility
functions.
   This code was written by Janel Hanmer September 2017 using SAS 9.4

   DATASET
   This code presumes the dataset's name is PROPrData

   INPUTS
   The input thetas must be a 7 element vector of the following form:
     - theta_cog is a score on the Cognitive Functioning - Abilities domain
     - theta_dep is a score on the Depression domain
     - theta_fat is a score on the Fatigue domain
     - theta_pain is a score on the Pain Interference domain
     - theta_phys is a score on the Physical Functioning domain
     - theta_slp is a score on the Sleep Disturbance domain
     - theta_sr is a score on the Ability to Participate in Social Roles and
Activities domain
   The thetas must be of the z-score form: usually a number from -3 to 3.
These are the scores constructed with a population mean of 0 and standard
deviation of 1.
   Note in particular, they should not be the "t-score" form.  These scores
are transformations of the z-scores such that the population mean is 50 and a
standard deviation of 10.

   OUTPUTS
   A number (utility) on the dead = 0, full health = 1 scale.
   Note that 1 is the maximum possible value, but scores less than 0 are
possible.
 */


 /*  The values we will need for the computation */

data PROPrData; set PROPrData;
  /* Coefficients for each turn points regression */

turncog1 =-2.052;
turncog2 =-1.565;
turncog3 =-1.239;
turncog4 =-0.902;
turncog5 =-0.649;
turncog6 =-0.367;
turncog7 =-0.002;
turncog8 =0.52;
turncog9 =1.124;

turndep1 =-1.082;
turndep2 =-0.264;
turndep3 =0.151;
turndep4 =0.596;
turndep5 =0.913;
```

```
turndep6 =1.388;
turndep7 =1.742;
turndep8 =2.245;
turndep9 =2.703;

turnfat1 =-1.648;
turnfat2 =-0.818;
turnfat3 =-0.094;
turnfat4 =0.303;
turnfat5 =0.87;
turnfat6 =1.124;
turnfat7 =1.688;
turnfat8 =2.053;
turnfat9 =2.423;

turnpain1 =-0.773;
turnpain2 =0.1;
turnpain3 =0.462;
turnpain4 =0.827;
turnpain5 =1.072;
turnpain6 =1.407;
turnpain7 =1.724;
turnpain8 =2.169;
turnpain9 =2.725;

turnphys1 =-2.575;
turnphys2 =-2.174;
turnphys3 =-1.784;
turnphys4 =-1.377;
turnphys5 =-0.787;
turnphys6 =-0.443;
turnphys7 =-0.211;
turnphys8 =0.16;
turnphys9 =0.966;

turnsleep1 =-1.535;
turnsleep2 =-0.775;
turnsleep3 =-0.459;
turnsleep4 =0.093;
turnsleep5 =0.335;
turnsleep6 =0.82;
turnsleep7 =1.659;
turnsleep8 =1.934;

turnsocial1 =-2.088;
turnsocial2 =-1.634;
turnsocial3 =-1.293;
turnsocial4 =-0.955;
turnsocial5 =-0.618;
turnsocial6 =-0.276;
turnsocial7 =0.083;
turnsocial8 =0.494;
turnsocial9 =1.221;

  /* Coefficients for each slope regression */
slopecog1 =-1.0047;
slopecog2 =-0.1745;
```

```
slopecog3 =-0.4223;
slopecog4 =-0.1949;
slopecog5 =-0.1082;
slopecog6 =-0.2468;
slopecog7 =-0.0176;
slopecog8 =-0.2192;


slopedep1 =0.1572;
slopedep2 =0;
slopedep3 =0.1793;
slopedep4 =0.1817;
slopedep5 =0.4109;
slopedep6 =0.1887;
slopedep7 =0.2115;
slopedep8 =0.7983;


slopefat1 =0.1152;
slopefat2 =0.1077;
slopefat3 =0.1189;
slopefat4 =0.1277;
slopefat5 =0.222;
slopefat6 =0.0496;
slopefat7 =0.3233;
slopefat8 =1.3632;


slopepain1 =0.0891;
slopepain2 =0.1721;
slopepain3 =0.1022;
slopepain4 =0.4241;
slopepain5 =0.3815;
slopepain6 =0.3681;
slopepain7 =0.1169;
slopepain8 =0.7594;


slopephys1 =-1.0761;
slopephys2 =-0.1756;
slopephys3 =-0.1764;
slopephys4 =-0.1161;
slopephys5 =-0.2721;
slopephys6 =-0.4082;
slopephys7 =-0.1695;
slopephys8 =-0.1346;


slopesleep1 =0.1241;
slopesleep2 =0;
slopesleep3 =0.0797;
slopesleep4 =0.3455;
slopesleep5 =0.3148;
slopesleep6 =0.1238;
slopesleep7 =1.8964;
```

```
slopesocial1 =-1.1152;
slopesocial2 =-0.2874;
slopesocial3 =-0.1352;
slopesocial4 =-0.132;
slopesocial5 =-0.4012;
slopesocial6 =0;
slopesocial7 =-0.054;
slopesocial8 =-0.201;


  /* Coefficients for each intercept points regression */
interceptcog1 =-1.0617;
interceptcog2 =0.2375;
interceptcog3 =-0.0694;
interceptcog4 =0.1357;
interceptcog5 =0.192;
interceptcog6 =0.1411;
interceptcog7 =0.1416;
interceptcog8 =0.2464;


interceptdep1 =0.1701;
interceptdep2 =0.1286;
interceptdep3 =0.1015;
interceptdep4 =0.1001;
interceptdep5 =-0.1092;
interceptdep6 =0.1993;
interceptdep7 =0.1595;
interceptdep8 =-1.1577;


interceptfat1 =0.1898;
interceptfat2 =0.1837;
interceptfat3 =0.1848;
interceptfat4 =0.1821;
interceptfat5 =0.1;
interceptfat6 =0.2938;
interceptfat7 =-0.1681;
interceptfat8 =-2.3031;


interceptpain1 =0.0689;
interceptpain2 =0.0606;
interceptpain3 =0.0929;
interceptpain4 =-0.1733;
interceptpain5 =-0.1277;
interceptpain6 =-0.1089;
interceptpain7 =0.3243;
interceptpain8 =-1.0692;


interceptphys1 =-1.7709;
interceptphys2 =0.1867;
interceptphys3 =0.1853;
interceptphys4 =0.2683;
interceptphys5 =0.1456;
interceptphys6 =0.0853;
```

```
interceptphys7 =0.1356;
interceptphys8 =0.13;


interceptsleep1 =0.1905;
interceptsleep2 =0.0943;
interceptsleep3 =0.1309;
interceptsleep4 =0.1062;
interceptsleep5 =0.1164;
interceptsleep6 =0.2731;
interceptsleep7 =-2.6676;


interceptsocial1 =-1.3285;
interceptsocial2 =0.0241;
interceptsocial3 =0.2209;
interceptsocial4 =0.2239;
interceptsocial5 =0.0576;
interceptsocial6 =0.1683;
interceptsocial7 =0.1728;
interceptsocial8 =0.2454;



 /*  Corner state disutilities*/
  c_cognition = 0.6350450;
  c_depression = 0.6661641;
  c_fatigue = 0.6386135;
  c_pain = 0.6529680;
  c_physical = 0.6883584;
  c_sleep = 0.5629657;
  c_social = 0.6112686;
  C = -0.9991828;

 /*  Constant for transforming from pits to dead */
  to_dead = 1.021915;


 /*  Create the output of each single-domain function*/

 /*  Cognition Disutility.  Higher cognition scores are better.*/
  cog_disutility = 1;
  if turncog1<=theta_cog<turncog2 then cog_disutility = interceptcog1 +
theta_cog * slopecog1;
  if turncog2<=theta_cog<turncog3 then cog_disutility = interceptcog2 +
theta_cog * slopecog2;
  if turncog3<=theta_cog<turncog4 then cog_disutility = interceptcog3 +
theta_cog * slopecog3;
  if turncog4<=theta_cog<turncog5 then cog_disutility = interceptcog4 +
theta_cog * slopecog4;
  if turncog5<=theta_cog<turncog6 then cog_disutility = interceptcog5 +
theta_cog * slopecog5;
  if turncog6<=theta_cog<turncog7 then cog_disutility = interceptcog6 +
theta_cog * slopecog6;
  if turncog7<=theta_cog<turncog8 then cog_disutility = interceptcog7 +
theta_cog * slopecog7;
```

```
  if turncog8<=theta_cog<turncog9 then cog_disutility = interceptcog8 +
theta_cog * slopecog8;
  if turncog9<=theta_cog then cog_disutility = 0;

 /*  Depression Disutility.  Lower depression scores are better*/
  dep_disutility=0;
  if turndep1<=theta_dep<turndep2 then dep_disutility = interceptdep1 +
theta_dep * slopedep1;
  if turndep2<=theta_dep<turndep3 then dep_disutility = interceptdep2 +
theta_dep * slopedep2;
  if turndep3<=theta_dep<turndep4 then dep_disutility = interceptdep3 +
theta_dep * slopedep3;
  if turndep4<=theta_dep<turndep5 then dep_disutility = interceptdep4 +
theta_dep * slopedep4;
  if turndep5<=theta_dep<turndep6 then dep_disutility = interceptdep5 +
theta_dep * slopedep5;
  if turndep6<=theta_dep<turndep7 then dep_disutility = interceptdep6 +
theta_dep * slopedep6;
  if turndep7<=theta_dep<turndep8 then dep_disutility = interceptdep7 +
theta_dep * slopedep7;
  if turndep8<=theta_dep<turndep9 then dep_disutility = interceptdep8 +
theta_dep * slopedep8;
  if turndep9<=theta_dep then dep_disutility = 1;


 /*  Fatigue Disutility.  Lower fatigue scores are better*/
  fat_disutility=0;
  if turnfat1<=theta_fat<turnfat2 then fat_disutility = interceptfat1 +
theta_fat * slopefat1;
  if turnfat2<=theta_fat<turnfat3 then fat_disutility = interceptfat2 +
theta_fat * slopefat2;
  if turnfat3<=theta_fat<turnfat4 then fat_disutility = interceptfat3 +
theta_fat * slopefat3;
  if turnfat4<=theta_fat<turnfat5 then fat_disutility = interceptfat4 +
theta_fat * slopefat4;
  if turnfat5<=theta_fat<turnfat6 then fat_disutility = interceptfat5 +
theta_fat * slopefat5;
  if turnfat6<=theta_fat<turnfat7 then fat_disutility = interceptfat6 +
theta_fat * slopefat6;
  if turnfat7<=theta_fat<turnfat8 then fat_disutility = interceptfat7 +
theta_fat * slopefat7;
  if turnfat8<=theta_fat<turnfat9 then fat_disutility = interceptfat8 +
theta_fat * slopefat8;
  if turnfat9<=theta_fat then fat_disutility = 1;


 /*  Pain Disutility.  Lower pain scores are better*/
  pain_disutility=0;
  if turnpain1<=theta_pain<turnpain2 then pain_disutility = interceptpain1 +
theta_pain * slopepain1;
  if turnpain2<=theta_pain<turnpain3 then pain_disutility = interceptpain2 +
theta_pain * slopepain2;
  if turnpain3<=theta_pain<turnpain4 then pain_disutility = interceptpain3 +
theta_pain * slopepain3;
  if turnpain4<=theta_pain<turnpain5 then pain_disutility = interceptpain4 +
theta_pain * slopepain4;
```

```
  if turnpain5<=theta_pain<turnpain6 then pain_disutility = interceptpain5 +
theta_pain * slopepain5;
  if turnpain6<=theta_pain<turnpain7 then pain_disutility = interceptpain6 +
theta_pain * slopepain6;
  if turnpain7<=theta_pain<turnpain8 then pain_disutility = interceptpain7 +
theta_pain * slopepain7;
  if turnpain8<=theta_pain<turnpain9 then pain_disutility = interceptpain8 +
theta_pain * slopepain8;
  if turnpain9<=theta_pain then pain_disutility = 1;


 /*  Physical Disutility.  higher physical function scores are better*/
  physical_disutility=1;
  if turnphys1<=theta_phys<turnphys2 then physical_disutility =
interceptphys1 + theta_phys * slopephys1;
  if turnphys2<=theta_phys<turnphys3 then physical_disutility =
interceptphys2 + theta_phys * slopephys2;
  if turnphys3<=theta_phys<turnphys4 then physical_disutility =
interceptphys3 + theta_phys * slopephys3;
  if turnphys4<=theta_phys<turnphys5 then physical_disutility =
interceptphys4 + theta_phys * slopephys4;
  if turnphys5<=theta_phys<turnphys6 then physical_disutility =
interceptphys5 + theta_phys * slopephys5;
  if turnphys6<=theta_phys<turnphys7 then physical_disutility =
interceptphys6 + theta_phys * slopephys6;
  if turnphys7<=theta_phys<turnphys8 then physical_disutility =
interceptphys7 + theta_phys * slopephys7;
  if turnphys8<=theta_phys<turnphys9 then physical_disutility =
interceptphys8 + theta_phys * slopephys8;
  if turnphys9<=theta_phys then physical_disutility = 0;


 /*  Sleep Disutility.  Lower sleep disturbance scores are better*/
  sleep_disutility=0;
  if turnsleep1<=theta_slp<turnsleep2 then sleep_disutility = interceptsleep1
+ theta_slp * slopesleep1;
  if turnsleep2<=theta_slp<turnsleep3 then sleep_disutility = interceptsleep2
+ theta_slp * slopesleep2;
  if turnsleep3<=theta_slp<turnsleep4 then sleep_disutility = interceptsleep3
+ theta_slp * slopesleep3;
  if turnsleep4<=theta_slp<turnsleep5 then sleep_disutility = interceptsleep4
+ theta_slp * slopesleep4;
  if turnsleep5<=theta_slp<turnsleep6 then sleep_disutility = interceptsleep5
+ theta_slp * slopesleep5;
  if turnsleep6<=theta_slp<turnsleep7 then sleep_disutility = interceptsleep6
+ theta_slp * slopesleep6;
  if turnsleep7<=theta_slp<turnsleep8 then sleep_disutility = interceptsleep7
+ theta_slp * slopesleep7;
  if turnsleep8<=theta_slp then sleep_disutility = 1;

 /*  Social Disutility.  Higher social scores are better*/
  social_disutility=1;
  if turnsocial1<=theta_sr<turnsocial2 then social_disutility =
interceptsocial1 + theta_sr * slopesocial1;
  if turnsocial2<=theta_sr<turnsocial3 then social_disutility =
interceptsocial2 + theta_sr * slopesocial2;
```

```
  if turnsocial3<=theta_sr<turnsocial4 then social_disutility =
interceptsocial3 + theta_sr * slopesocial3;
  if turnsocial4<=theta_sr<turnsocial5 then social_disutility =
interceptsocial4 + theta_sr * slopesocial4;
  if turnsocial5<=theta_sr<turnsocial6 then social_disutility =
interceptsocial5 + theta_sr * slopesocial5;
  if turnsocial6<=theta_sr<turnsocial7 then social_disutility =
interceptsocial6 + theta_sr * slopesocial6;
  if turnsocial7<=theta_sr<turnsocial8 then social_disutility =
interceptsocial7 + theta_sr * slopesocial7;
  if turnsocial8<=theta_sr<turnsocial9 then social_disutility =
interceptsocial8 + theta_sr * slopesocial8;
  if turnsocial9<=theta_sr then social_disutility = 0;


 /*  Now, plug it into the multiattribute disutility function */

  multi_attribute_disutility = (1/C) * ((1 + C * c_cognition *
cog_disutility)* (1 + C * c_depression * dep_disutility)*
 (1 + C * c_fatigue * fat_disutility)* (1 + C * c_pain * pain_disutility)*
 (1 + C * c_physical * physical_disutility)* (1 + C * c_sleep *
sleep_disutility)* (1 + C * c_social * social_disutility) - 1);


  /* Now make it a utility, on the dead/full health scale */
  PROPr = round (1 - to_dead * multi_attribute_disutility, 0.001);
run;

  /* single attribute utility functions */

data PROPrData; set PROPrData;
      cognition_utility = round (1 - cog_disutility, 0.001);
      depression_utility = round (1 - dep_disutility, 0.001);
      fatigue_utility = round (1 - fat_disutility, 0.001);
      pain_utility = round (1 - pain_disutility, 0.001);
      physical_utility = round (1 - physical_disutility, 0.001);
      sleep_utility = round (1 - sleep_disutility, 0.001);
      social_utility = round (1 - social_disutility, 0.001); run;


/* clean up dataset */
data PROPrData; set PROPrData;
 drop multi_attribute_disutility  to_dead
 turncog1   turncog2    turncog3    turncog4    turncog5    turncog6
      turncog7    turncog8    turncog9          turndep1    turndep2
      turndep3    turndep4    turndep5    turndep6    turndep7    turndep8
      turndep9          turnfat1    turnfat2    turnfat3    turnfat4
      turnfat5    turnfat6    turnfat7    turnfat8    turnfat9
      turnpain1   turnpain2   turnpain3   turnpain4   turnpain5   turnpain6
      turnpain7   turnpain8   turnpain9         turnphys1   turnphys2
      turnphys3   turnphys4   turnphys5   turnphys6   turnphys7   turnphys8
      turnphys9         turnsleep1  turnsleep2  turnsleep3  turnsleep4
      turnsleep5  turnsleep6  turnsleep7  turnsleep8        turnsocial1
      turnsocial2       turnsocial3       turnsocial4       turnsocial5
      turnsocial6       turnsocial7       turnsocial8       turnsocial9
slopecog1   slopecog2   slopecog3   slopecog4   slopecog5   slopecog6
      slopecog7   slopecog8               slopedep1   slopedep2   slopedep3
```

```
      slopedep4   slopedep5   slopedep6   slopedep7   slopedep8
      slopefat1   slopefat2   slopefat3   slopefat4   slopefat5   slopefat6
      slopefat7   slopefat8               slopepain1  slopepain2  slopepain3
      slopepain4  slopepain5  slopepain6  slopepain7  slopepain8
      slopephys1  slopephys2  slopephys3  slopephys4  slopephys5  slopephys6
      slopephys7  slopephys8              slopesleep1       slopesleep2
      slopesleep3       slopesleep4       slopesleep5       slopesleep6
      slopesleep7                   slopesocial1      slopesocial2
      slopesocial3      slopesocial4      slopesocial5      slopesocial6
      slopesocial7      slopesocial8
interceptcog1     interceptcog2     interceptcog3     interceptcog4
      interceptcog5     interceptcog6     interceptcog7     interceptcog8
      interceptdep1     interceptdep2     interceptdep3     interceptdep4
      interceptdep5     interceptdep6     interceptdep7     interceptdep8
            interceptfat1     interceptfat2     interceptfat3
      interceptfat4     interceptfat5     interceptfat6     interceptfat7
      interceptfat8                 interceptpain1    interceptpain2
      interceptpain3    interceptpain4    interceptpain5    interceptpain6
      interceptpain7    interceptpain8                interceptphys1
      interceptphys2    interceptphys3    interceptphys4    interceptphys5
      interceptphys6    interceptphys7    interceptphys8
      interceptsleep1   interceptsleep2   interceptsleep3   interceptsleep4
      interceptsleep5   interceptsleep6   interceptsleep7
      interceptsocial1  interceptsocial2  interceptsocial3  interceptsocial4
      interceptsocial5  interceptsocial6  interceptsocial7  interceptsocial8
c_cognition c_depression c_fatigue c_pain c_physical c_sleep c_social c
cog_disutility dep_disutility fat_disutility pain_disutility
physical_disutility sleep_disutility social_disutility; run;
```

Appendix 2: R Code

Below, we present standalone R code for the multi-attribute PROPr scoring function, along with the single-attribute functions. As input, the function takes a vector of theta scores, arranged as described in the code comments. The function outputs a list of utilities: first, the utility produced from the multi-attribute scoring function, and then the seven single-attribute scores. Note that the utility produced from the multi-attribute scoring function is on a scale where 0 is the utility of dead, 1 is the utility of full health, 1 is the highest utility, and utilities lower than 0 are possible. The seven single-attribute functions are on seven different scales where 0 is the utility of the worst description in the given domain, and 1 is the utility of the best description in the given domain—and all utilities are between 0 and 1.

```
propr.maut.function.201709 <- function(thetas){

  # DESCRIPTION This is the multi-attribute utility function, using isotonic
  # regression with linear interpolation for the single-attribute (dis)utility
  # functions. This code was written by Barry Dewitt in September 2017 using R
  # 3.4.0.

  # INPUTS
  # The input thetas must be a 7 element vector with the following components
  # (in order):
  #   - theta_cog is a score on the Cognitive Functioning - Abilities domain
  #   - theta_dep is a score on the Depression domain
  #   - theta_fat is a score on the Fatigue domain
  #   - theta_pain is a score on the Pain Interference domain
  #   - theta_phys is a score on the Physical Functioning domain
  #   - theta_slp is a score on the Sleep Disturbance domain
  #   - theta_sr is a score on the Ability to Participate in Social Roles
  # and Activities domain

  # The thetas must be of the z-score form: usually a number from -3 to 3.
  # These are the scores constructed with a population mean of 0 and standard
  # deviation of 1. Note in particular, they should not be the "t-score" form.
  # These scores are transformations of the z-scores such that the population
  # mean is 50 and a standard deviation of 10.

  # OUTPUTS
  # A list with the following components:

  # PROPr -- A number (utility) on the dead = 0, full health = 1 scale. (Note
  # that 1 is the maximum possible value, but scores less than 0 are possible.)
  # One single-attribute utility score for each domain, where the utility of
  # the (disutility) corner state = 0, and full health = 1. (Note scores are
  # bounded by 0 and 1 for the single-attribute scales.) These components are
  # labeled by the domain names.

  # Label input components
  theta_cog <- thetas[1]
  theta_dep <- thetas[2]
  theta_fat <- thetas[3]
  theta_pain <- thetas[4]
  theta_phys <- thetas[5]
  theta_slp <- thetas[6]
  theta_sr <- thetas[7]

  # Values where the line segments of the isotonic regression with interpolation
  # change
```

```
turncog1 <- -2.052
turncog2 <- -1.565
turncog3 <- -1.239
turncog4 <- -0.902
turncog5 <- -0.649
turncog6 <- -0.367
turncog7 <- -0.002
turncog8 <- 0.52
turncog9 <- 1.124

turndep1 <- -1.082
turndep2 <- -0.264
turndep3 <- 0.151
turndep4 <- 0.596
turndep5 <- 0.913
turndep6 <- 1.388
turndep7 <- 1.742
turndep8 <- 2.245
turndep9 <- 2.703

turnfat1 <- -1.648
turnfat2 <- -0.818
turnfat3 <- -0.094
turnfat4 <- 0.303
turnfat5 <- 0.87
turnfat6 <- 1.124
turnfat7 <- 1.688
turnfat8 <- 2.053
turnfat9 <- 2.423

turnpain1 <- -0.773
turnpain2 <- 0.1
turnpain3 <- 0.462
turnpain4 <- 0.827
turnpain5 <- 1.072
turnpain6 <- 1.407
turnpain7 <- 1.724
turnpain8 <- 2.169
turnpain9 <- 2.725

turnphys1 <- -2.575
turnphys2 <- -2.174
turnphys3 <- -1.784
turnphys4 <- -1.377
turnphys5 <- -0.787
turnphys6 <- -0.443
turnphys7 <- -0.211
turnphys8 <- 0.16
turnphys9 <- 0.966

turnsleep1 <- -1.535
turnsleep2 <- -0.775
turnsleep3 <- -0.459
turnsleep4 <- 0.093
turnsleep5 <- 0.335
turnsleep6 <- 0.82
turnsleep7 <- 1.659
turnsleep8 <- 1.934

turnsocial1 <- -2.088
turnsocial2 <- -1.634
turnsocial3 <- -1.293
turnsocial4 <- -0.955
```

```
turnsocial5 <- -0.618
turnsocial6 <- -0.276
turnsocial7 <- 0.083
turnsocial8 <- 0.494
turnsocial9 <- 1.221

# Slopes of each line segment specification

slopecog1 <- -1.0047
slopecog2 <- -0.1745
slopecog3 <- -0.4223
slopecog4 <- -0.1949
slopecog5 <- -0.1082
slopecog6 <- -0.2468
slopecog7 <- -0.0176
slopecog8 <- -0.2192


slopedep1 <- 0.1572
slopedep2 <- 0
slopedep3 <- 0.1793
slopedep4 <- 0.1817
slopedep5 <- 0.4109
slopedep6 <- 0.1887
slopedep7 <- 0.2115
slopedep8 <- 0.7983


slopefat1 <- 0.1152
slopefat2 <- 0.1077
slopefat3 <- 0.1189
slopefat4 <- 0.1277
slopefat5 <- 0.222
slopefat6 <- 0.0496
slopefat7 <- 0.3233
slopefat8 <- 1.3632


slopepain1 <- 0.0891
slopepain2 <- 0.1721
slopepain3 <- 0.1022
slopepain4 <- 0.4241
slopepain5 <- 0.3815
slopepain6 <- 0.3681
slopepain7 <- 0.1169
slopepain8 <- 0.7594


slopephys1 <- -1.0761
slopephys2 <- -0.1756
slopephys3 <- -0.1764
slopephys4 <- -0.1161
slopephys5 <- -0.2721
slopephys6 <- -0.4082
slopephys7 <- -0.1695
slopephys8 <- -0.1346


slopesleep1 <- 0.1241
slopesleep2 <- 0
slopesleep3 <- 0.0797
slopesleep4 <- 0.3455
slopesleep5 <- 0.3148
```

```
slopesleep6 <- 0.1238
slopesleep7 <- 1.8964


slopesocial1 <- -1.1152
slopesocial2 <- -0.2874
slopesocial3 <- -0.1352
slopesocial4 <- -0.132
slopesocial5 <- -0.4012
slopesocial6 <- 0
slopesocial7 <- -0.054
slopesocial8 <- -0.201


# Intercepts of each line segment specification

interceptcog1 <- -1.0617
interceptcog2 <- 0.2375
interceptcog3 <- -0.0694
interceptcog4 <- 0.1357
interceptcog5 <- 0.192
interceptcog6 <- 0.1411
interceptcog7 <- 0.1416
interceptcog8 <- 0.2464


interceptdep1 <- 0.1701
interceptdep2 <- 0.1286
interceptdep3 <- 0.1015
interceptdep4 <- 0.1001
interceptdep5 <- -0.1092
interceptdep6 <- 0.1993
interceptdep7 <- 0.1595
interceptdep8 <- -1.1577


interceptfat1 <- 0.1898
interceptfat2 <- 0.1837
interceptfat3 <- 0.1848
interceptfat4 <- 0.1821
interceptfat5 <- 0.1
interceptfat6 <- 0.2938
interceptfat7 <- -0.1681
interceptfat8 <- -2.3031


interceptpain1 <- 0.0689
interceptpain2 <- 0.0606
interceptpain3 <- 0.0929
interceptpain4 <- -0.1733
interceptpain5 <- -0.1277
interceptpain6 <- -0.1089
interceptpain7 <- 0.3243
interceptpain8 <- -1.0692


interceptphys1 <- -1.7709
interceptphys2 <- 0.1867
interceptphys3 <- 0.1853
interceptphys4 <- 0.2683
interceptphys5 <- 0.1456
interceptphys6 <- 0.0853
interceptphys7 <- 0.1356
```

```
interceptphys8 <- 0.13


interceptsleep1 <- 0.1905
interceptsleep2 <- 0.0943
interceptsleep3 <- 0.1309
interceptsleep4 <- 0.1062
interceptsleep5 <- 0.1164
interceptsleep6 <- 0.2731
interceptsleep7 <- -2.6676


interceptsocial1 <- -1.3285
interceptsocial2 <- 0.0241
interceptsocial3 <- 0.2209
interceptsocial4 <- 0.2239
interceptsocial5 <- 0.0576
interceptsocial6 <- 0.1683
interceptsocial7 <- 0.1728
interceptsocial8 <- 0.2454


# Corner state disutility values
c_cognition <-  0.6350450
c_depression <-  0.6661641
c_fatigue <-  0.6386135
c_pain <-  0.6529680
c_physical <-  0.6883584
c_sleep <-  0.5629657
c_social <-  0.6112686
C <-  -0.9991828

# Constant for transforming from all-worst = 0 to dead = 0
to_dead <-  1.021915

# Create output of each single-attribute disutility function

# Cognition Disutility.  Higher cognition scores are better.
cog_disutility <- 1
if(turncog1<=theta_cog & theta_cog <turncog2) {cog_disutility <-
  interceptcog1 + theta_cog * slopecog1}
if(turncog2<=theta_cog & theta_cog <turncog3) {cog_disutility <-
  interceptcog2 + theta_cog * slopecog2}
if(turncog3<=theta_cog & theta_cog <turncog4) {cog_disutility <-
  interceptcog3 + theta_cog * slopecog3}
if(turncog4<=theta_cog & theta_cog <turncog5) {cog_disutility <-
  interceptcog4 + theta_cog * slopecog4}
if(turncog5<=theta_cog & theta_cog <turncog6) {cog_disutility <-
  interceptcog5 + theta_cog * slopecog5}
if(turncog6<=theta_cog & theta_cog <turncog7) {cog_disutility <-
  interceptcog6 + theta_cog * slopecog6}
if(turncog7<=theta_cog & theta_cog <turncog8) {cog_disutility <-
  interceptcog7 + theta_cog * slopecog7}
if(turncog8<=theta_cog & theta_cog <turncog9) {cog_disutility <-
  interceptcog8 + theta_cog * slopecog8}
if(turncog9<=theta_cog) {cog_disutility <- 0}

# Depression Disutility.  Lower depression scores are better
dep_disutility <- 0
if(turndep1<=theta_dep & theta_dep <turndep2) {dep_disutility <-
  interceptdep1 + theta_dep * slopedep1}
if(turndep2<=theta_dep & theta_dep <turndep3) {dep_disutility <-
  interceptdep2 + theta_dep * slopedep2}
```

```
if(turndep3<=theta_dep & theta_dep <turndep4) {dep_disutility <-
  interceptdep3 + theta_dep * slopedep3}
if(turndep4<=theta_dep & theta_dep <turndep5) {dep_disutility <-
  interceptdep4 + theta_dep * slopedep4}
if(turndep5<=theta_dep & theta_dep <turndep6) {dep_disutility <-
  interceptdep5 + theta_dep * slopedep5}
if(turndep6<=theta_dep & theta_dep <turndep7) {dep_disutility <-
  interceptdep6 + theta_dep * slopedep6}
if(turndep7<=theta_dep & theta_dep <turndep8) {dep_disutility <-
  interceptdep7 + theta_dep * slopedep7}
if(turndep8<=theta_dep & theta_dep <turndep9) {dep_disutility <-
  interceptdep8 + theta_dep * slopedep8}
if(turndep9<=theta_dep) {dep_disutility <- 1}


# Fatigue Disutility.  Lower fatigue scores are better
fat_disutility <- 0
if(turnfat1<=theta_fat & theta_fat <turnfat2) {fat_disutility <-
  interceptfat1 + theta_fat * slopefat1}
if(turnfat2<=theta_fat & theta_fat <turnfat3) {fat_disutility <-
  interceptfat2 + theta_fat * slopefat2}
if(turnfat3<=theta_fat & theta_fat <turnfat4) {fat_disutility <-
  interceptfat3 + theta_fat * slopefat3}
if(turnfat4<=theta_fat & theta_fat <turnfat5) {fat_disutility <-
  interceptfat4 + theta_fat * slopefat4}
if(turnfat5<=theta_fat & theta_fat <turnfat6) {fat_disutility <-
  interceptfat5 + theta_fat * slopefat5}
if(turnfat6<=theta_fat & theta_fat <turnfat7) {fat_disutility <-
  interceptfat6 + theta_fat * slopefat6}
if(turnfat7<=theta_fat & theta_fat <turnfat8) {fat_disutility <-
  interceptfat7 + theta_fat * slopefat7}
if(turnfat8<=theta_fat & theta_fat <turnfat9) {fat_disutility <-
  interceptfat8 + theta_fat * slopefat8}
if(turnfat9<=theta_fat) {fat_disutility <- 1}


#  Pain Disutility.  Lower pain scores are better
pain_disutility <- 0
if(turnpain1<=theta_pain & theta_pain <turnpain2) {pain_disutility <-
  interceptpain1 + theta_pain * slopepain1}
if(turnpain2<=theta_pain & theta_pain <turnpain3) {pain_disutility <-
  interceptpain2 + theta_pain * slopepain2}
if(turnpain3<=theta_pain & theta_pain <turnpain4) {pain_disutility <-
  interceptpain3 + theta_pain * slopepain3}
if(turnpain4<=theta_pain & theta_pain <turnpain5) {pain_disutility <-
  interceptpain4 + theta_pain * slopepain4}
if(turnpain5<=theta_pain & theta_pain <turnpain6) {pain_disutility <-
  interceptpain5 + theta_pain * slopepain5}
if(turnpain6<=theta_pain & theta_pain <turnpain7) {pain_disutility <-
  interceptpain6 + theta_pain * slopepain6}
if(turnpain7<=theta_pain & theta_pain <turnpain8) {pain_disutility <-
  interceptpain7 + theta_pain * slopepain7}
if(turnpain8<=theta_pain & theta_pain <turnpain9) {pain_disutility <-
  interceptpain8 + theta_pain * slopepain8}
if(turnpain9<=theta_pain) {pain_disutility <- 1}


#  Physical Disutility.  Higher physical function scores are better
physical_disutility <- 1
if(turnphys1<=theta_phys & theta_phys <turnphys2) {physical_disutility <-
  interceptphys1 + theta_phys * slopephys1}
if(turnphys2<=theta_phys & theta_phys <turnphys3) {physical_disutility <-
  interceptphys2 + theta_phys * slopephys2}
```

```
if(turnphys3<=theta_phys & theta_phys <turnphys4) {physical_disutility <-
  interceptphys3 + theta_phys * slopephys3}
if(turnphys4<=theta_phys & theta_phys <turnphys5) {physical_disutility <-
  interceptphys4 + theta_phys * slopephys4}
if(turnphys5<=theta_phys & theta_phys <turnphys6) {physical_disutility <-
  interceptphys5 + theta_phys * slopephys5}
if(turnphys6<=theta_phys & theta_phys <turnphys7) {physical_disutility <-
  interceptphys6 + theta_phys * slopephys6}
if(turnphys7<=theta_phys & theta_phys <turnphys8) {physical_disutility <-
  interceptphys7 + theta_phys * slopephys7}
if(turnphys8<=theta_phys & theta_phys <turnphys9) {physical_disutility <-
  interceptphys8 + theta_phys * slopephys8}
if(turnphys9<=theta_phys) {physical_disutility <- 0}


#  Sleep Disutility.  Lower sleep disturbance scores are better
sleep_disutility <- 0
if(turnsleep1<=theta_slp & theta_slp <turnsleep2) {sleep_disutility <-
  interceptsleep1 + theta_slp * slopesleep1}
if(turnsleep2<=theta_slp & theta_slp <turnsleep3) {sleep_disutility <-
  interceptsleep2 + theta_slp * slopesleep2}
if(turnsleep3<=theta_slp & theta_slp <turnsleep4) {sleep_disutility <-
  interceptsleep3 + theta_slp * slopesleep3}
if(turnsleep4<=theta_slp & theta_slp <turnsleep5) {sleep_disutility <-
  interceptsleep4 + theta_slp * slopesleep4}
if(turnsleep5<=theta_slp & theta_slp <turnsleep6) {sleep_disutility <-
  interceptsleep5 + theta_slp * slopesleep5}
if(turnsleep6<=theta_slp & theta_slp <turnsleep7) {sleep_disutility <-
  interceptsleep6 + theta_slp * slopesleep6}
if(turnsleep7<=theta_slp & theta_slp <turnsleep8) {sleep_disutility <-
  interceptsleep7 + theta_slp * slopesleep7}
if(turnsleep8<=theta_slp) {sleep_disutility <- 1}

#  Social Disutility.  Higher social scores are better
social_disutility <- 1
if(turnsocial1<=theta_sr & theta_sr <turnsocial2) {social_disutility <-
  interceptsocial1 + theta_sr * slopesocial1}
if(turnsocial2<=theta_sr & theta_sr <turnsocial3) {social_disutility <-
  interceptsocial2 + theta_sr * slopesocial2}
if(turnsocial3<=theta_sr & theta_sr <turnsocial4) {social_disutility <-
  interceptsocial3 + theta_sr * slopesocial3}
if(turnsocial4<=theta_sr & theta_sr <turnsocial5) {social_disutility <-
  interceptsocial4 + theta_sr * slopesocial4}
if(turnsocial5<=theta_sr & theta_sr <turnsocial6) {social_disutility <-
  interceptsocial5 + theta_sr * slopesocial5}
if(turnsocial6<=theta_sr & theta_sr <turnsocial7) {social_disutility <-
  interceptsocial6 + theta_sr * slopesocial6}
if(turnsocial7<=theta_sr & theta_sr <turnsocial8) {social_disutility <-
  interceptsocial7 + theta_sr * slopesocial7}
if(turnsocial8<=theta_sr & theta_sr <turnsocial9) {social_disutility <-
  interceptsocial8 + theta_sr * slopesocial8}
if(turnsocial9<=theta_sr) {social_disutility <- 0}


# Now, plug it into the multiattribute disutility function

multi_attribute_disutility <-
  (1/C) * ((1 + C * c_cognition * cog_disutility)*
           (1 + C * c_depression * dep_disutility)*
           (1 + C * c_fatigue * fat_disutility)*
           (1 + C * c_pain * pain_disutility)*
           (1 + C * c_physical * physical_disutility)*
           (1 + C * c_sleep * sleep_disutility)*
```

```
              (1 + C * c_social * social_disutility) - 1)

  # Now make it a utility, on the dead/full health scale
  PROPr <- round(1 - to_dead * multi_attribute_disutility, 3)

  # Single attribute utility functions
  cognition_utility <- round(1 - cog_disutility, 3)
  depression_utility <- round(1 - dep_disutility, 3)
  fatigue_utility <- round(1 - fat_disutility, 3)
  pain_utility <- round(1 - pain_disutility, 3)
  physical_utility <- round(1 - physical_disutility, 3)
  sleep_utility <- round(1 - sleep_disutility, 3)
  social_utility <- round(1 - social_disutility, 3)


  # Return PROPr multi-attribute score on dead-full health scale, and individual
  # scores on each of the single-attribute functions, where 0 = disutility
  # corner state and 1 = full health.
  propr.values <- list(PROPr = PROPr,
                     cognition = cognition_utility,
                     depression = depression_utility,
                     fatigue = fatigue_utility,
                     pain = pain_utility,
                     physical = physical_utility,
                     sleep = sleep_utility,
                     social = social_utility)

  return(propr.values)
}
```

Appendix 3: Health State Item Combinations

Cognition

Cognition	I have been able to concentrate. . .	Not at all	A little bit	Somewhat	**Quite a bit**	Very much
	I have been able to remember to do things, like take medicine or buy something I needed . . .	Not at all	A little bit	Somewhat	Quite a bit	**Very much**

Cognition	I have been able to concentrate. . .	Not at all	A little bit	Somewhat	**Quite a bit**	Very much
	I have been able to remember to do things, like take medicine or buy something I needed . . .	Not at all	A little bit	Somewhat	**Quite a bit**	Very much

Cognition	I have been able to concentrate. . .	Not at all	A little bit	Somewhat	**Quite a bit**	Very much
	I have been able to remember to do things, like take medicine or buy something I needed . . .	Not at all	A little bit	**Somewhat**	Quite a bit	Very much

Cognition	I have been able to concentrate. . .	Not at all	A little bit	**Somewhat**	Quite a bit	Very much
	I have been able to remember to do things, like take medicine or buy something I needed . . .	Not at all	A little bit	**Somewhat**	Quite a bit	Very much

Cognition	I have been able to concentrate. . .	Not at all	**A little bit**	Somewhat	Quite a bit	Very much
	I have been able to remember to do things, like take medicine or buy something I needed . . .	Not at all	A little bit	**Somewhat**	Quite a bit	Very much

Cognition	I have been able to concentrate. . .	Not at all	**A little bit**	Somewhat	Quite a bit	Very much
	I have been able to remember to do things, like take medicine or buy something I needed . . .	Not at all	**A little bit**	Somewhat	Quite a bit	Very much

Cognition	I have been able to concentrate. . .	**Not at all**	A little bit	Somewhat	Quite a bit	Very much
	I have been able to remember to do things, like take medicine or buy something I needed . . .	Not at all	**A little bit**	Somewhat	Quite a bit	Very much

Cognition	I have been able to concentrate. . .	**Not at all**	A little bit	Somewhat	Quite a bit	Very much
	I have been able to remember to do things, like take medicine or buy something I needed . . .	**Not at all**	A little bit	Somewhat	Quite a bit	Very much

Depression

Dep	I felt unhappy . . .	Always	Often	Sometimes	**Rarely**	Never
	I felt that nothing was interesting . . .	Always	Often	Sometimes	Rarely	**Never**

Dep	I felt unhappy . . .	Always	Often	Sometimes	**Rarely**	Never
	I felt that nothing was interesting . . .	Always	Often	Sometimes	**Rarely**	Never

Dep	I felt unhappy . . .	Always	Often	**Sometimes**	Rarely	Never
	I felt that nothing was interesting . . .	Always	Often	Sometimes	**Rarely**	Never

Dep	I felt unhappy . . .	Always	Often	**Sometimes**	Rarely	Never
	I felt that nothing was interesting . . .	Always	Often	**Sometimes**	Rarely	Never

Dep	I felt unhappy . . .	Always	**Often**	Sometimes	Rarely	Never
	I felt that nothing was interesting . . .	Always	Often	**Sometimes**	Rarely	Never

Dep	I felt unhappy . . .	Always	**Often**	Sometimes	Rarely	Never
	I felt that nothing was interesting . . .	Always	**Often**	Sometimes	Rarely	Never

Dep	I felt unhappy . . .	**Always**	Often	Sometimes	Rarely	Never
	I felt that nothing was interesting . . .	Always	**Often**	Sometimes	Rarely	Never

Dep	I felt unhappy . . .	**Always**	Often	Sometimes	Rarely	Never
	I felt that nothing was interesting . . .	**Always**	Often	Sometimes	Rarely	Never

Fatigue

Fatigue	How often were you too tired to take a bath or shower? . . .	Always	Often	Sometimes	Rarely	**Never**
	How often did you feel tired?	Always	Often	Sometimes	**Rarely**	Never

Fatigue	How often were you too tired to take a bath or shower? . . .	Always	Often	Sometimes	Rarely	**Never**
	How often did you feel tired?	Always	Often	**Sometimes**	Rarely	Never

Fatigue	How often were you too tired to take a bath or shower? . . .	Always	Often	Sometimes	**Rarely**	Never
	How often did you feel tired?	Always	Often	**Sometimes**	Rarely	Never

Fatigue	How often were you too tired to take a bath or shower? . . .	Always	Often	Sometimes	**Rarely**	Never
	How often did you feel tired?	Always	**Often**	Sometimes	Rarely	Never

Fatigue	How often were you too tired to take a bath or shower? . . .	Always	Often	**Sometimes**	Rarely	Never
	How often did you feel tired?	Always	**Often**	Sometimes	Rarely	Never

Fatigue	How often were you too tired to take a bath or shower? . . .	Always	Often	**Sometimes**	Rarely	Never
	How often did you feel tired?	**Always**	Often	Sometimes	Rarely	Never

Fatigue	How often were you too tired to take a bath or shower? . . .	Always	**Often**	Sometimes	Rarely	Never
	How often did you feel tired?	**Always**	Often	Sometimes	Rarely	Never

Fatigue	How often were you too tired to take a bath or shower? . . .	**Always**	Often	Sometimes	Rarely	Never
	How often did you feel tired?	**Always**	Often	Sometimes	Rarely	Never

Pain

Pain	How often was your pain so severe you could think of nothing else? . . .	Always	Often	Sometimes	Rarely	**Never**
	How often was pain distressing to you?. . .	Always	Often	Sometimes	**Rarely**	Never

Pain	How often was your pain so severe you could think of nothing else? . . .	Always	Often	Sometimes	Rarely	**Never**
	How often was pain distressing to you?. . .	Always	Often	**Sometimes**	Rarely	Never

Pain	How often was your pain so severe you could think of nothing else? . . .	Always	Often	Sometimes	**Rarely**	Never
	How often was pain distressing to you?. . .	Always	Often	**Sometimes**	Rarely	Never

Pain	How often was your pain so severe you could think of nothing else? . . .	Always	Often	**Sometimes**	Rarely	Never
	How often was pain distressing to you?. . .	Always	Often	**Sometimes**	Rarely	Never

Pain	How often was your pain so severe you could think of nothing else? . . .	Always	Often	**Sometimes**	Rarely	Never
	How often was pain distressing to you?. . .	Always	**Often**	Sometimes	Rarely	Never

Pain	How often was your pain so severe you could think of nothing else? . . .	Always	**Often**	Sometimes	Rarely	Never
	How often was pain distressing to you?. . .	Always	**Often**	Sometimes	Rarely	Never

Pain	How often was your pain so severe you could think of nothing else? . . .	Always	**Often**	Sometimes	Rarely	Never
	How often was pain distressing to you?. . .	**Always**	Often	Sometimes	Rarely	Never

Pain	How often was your pain so severe you could think of nothing else? . . .	**Always**	Often	Sometimes	Rarely	Never
	How often was pain distressing to you?. . .	**Always**	Often	Sometimes	Rarely	Never

Physical Function

Physical	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	**Without any difficulty**
	Are you able to run 100 yards (100 m)? . . .	Unable to do	With much difficulty	With some difficulty	**With a little difficulty**	Without any difficulty

Physical	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	**Without any difficulty**
	Are you able to run 100 yards (100 m)? . . .	Unable to do	With much difficulty	**With some difficulty**	With a little difficulty	Without any difficulty

Physical	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	**Without any difficulty**
	Are you able to run 100 yards (100 m)? . . .	Unable to do	**With much difficulty**	With some difficulty	With a little difficulty	Without any difficulty

Physical	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	Unable to do	With much difficulty	With some difficulty	With a little difficulty	**Without any difficulty**
	Are you able to run 100 yards (100 m)? . . .	**Unable to do**	With much difficulty	With some difficulty	With a little difficulty	Without any difficulty

Physical	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	Unable to do	With much difficulty	With some difficulty	**With a little difficulty**	Without any difficulty
	Are you able to run 100 yards (100 m)? . . .	**Unable to do**	With much difficulty	With some difficulty	With a little difficulty	Without any difficulty

Physical	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	Unable to do	With much difficulty	**With some difficulty**	With a little difficulty	Without any difficulty
	Are you able to run 100 yards (100 m)? . . .	**Unable to do**	With much difficulty	With some difficulty	With a little difficulty	Without any difficulty

Physical	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	Unable to do	**With much difficulty**	With some difficulty	With a little difficulty	Without any difficulty
	Are you able to run 100 yards (100 m)? . . .	**Unable to do**	With much difficulty	With some difficulty	With a little difficulty	Without any difficulty

Physical	Are you able to dress yourself, including tying shoelaces and buttoning up your clothes? . . .	**Unable to do**	With much difficulty	With some difficulty	With a little difficulty	Without any difficulty
	Are you able to run 100 yards (100 m)? . . .	**Unable to do**	With much difficulty	With some difficulty	With a little difficulty	Without any difficulty

Sleep Disturbance

Sleep	I got enough sleep . . .	Never	Rarely	Sometimes	**Often**	Always
	I woke up too early and could not fall back to sleep . . .	Always	Often	Sometimes	Rarely	**Never**

Sleep	I got enough sleep . . .	Never	Rarely	Sometimes	**Often**	Always
	I woke up too early and could not fall back to sleep . . .	Always	Often	Sometimes	**Rarely**	Never

Sleep	I got enough sleep . . .	Never	Rarely	**Sometimes**	Often	Always
	I woke up too early and could not fall back to sleep . . .	Always	Often	Sometimes	**Rarely**	Never

Sleep	I got enough sleep . . .	Never	Rarely	**Sometimes**	Often	Always
	I woke up too early and could not fall back to sleep . . .	Always	Often	**Sometimes**	Rarely	Never

Sleep	I got enough sleep . . .	Never	**Rarely**	Sometimes	Often	Always
	I woke up too early and could not fall back to sleep . . .	Always	Often	**Sometimes**	Rarely	Never

Sleep	I got enough sleep . . .	**Never**	Rarely	Sometimes	Often	Always
	I woke up too early and could not fall back to sleep . . .	Always	**Often**	Sometimes	Rarely	Never

Sleep	I got enough sleep . . .	**Never**	Rarely	Sometimes	Often	Always
	I woke up too early and could not fall back to sleep . . .	**Always**	Often	Sometimes	Rarely	Never

Social Roles

Social	I have trouble taking care of my regular personal responsibilities . . .	Always	Usually	Sometimes	Rarely	**Never**
	I have trouble participating in recreational activities with others. . .	Always	Usually	Sometimes	**Rarely**	Never

Social	I have trouble taking care of my regular personal responsibilities . . .	Always	Usually	Sometimes	**Rarely**	Never
	I have trouble participating in recreational activities with others. . .	Always	Usually	Sometimes	**Rarely**	Never

Social	I have trouble taking care of my regular personal responsibilities . . .	Always	Usually	Sometimes	**Rarely**	Never
	I have trouble participating in recreational activities with others. . .	Always	Usually	**Sometimes**	Rarely	Never

Social	I have trouble taking care of my regular personal responsibilities . . .	Always	Usually	**Sometimes**	Rarely	Never
	I have trouble participating in recreational activities with others. . .	Always	Usually	**Sometimes**	Rarely	Never

Social	I have trouble taking care of my regular personal responsibilities . . .	Always	Usually	**Sometimes**	Rarely	Never
	I have trouble participating in recreational activities with others. . .	Always	**Usually**	Sometimes	Rarely	Never

Social	I have trouble taking care of my regular personal responsibilities . . .	Always	Usually	**Sometimes**	Rarely	Never
	I have trouble participating in recreational activities with others. . .	**Always**	Usually	Sometimes	Rarely	Never

Social	I have trouble taking care of my regular personal responsibilities . . .	Always	**Usually**	Sometimes	Rarely	Never
	I have trouble participating in recreational activities with others. . .	**Always**	Usually	Sometimes	Rarely	Never

Social	I have trouble taking care of my regular personal responsibilities . . .	**Always**	Usually	Sometimes	Rarely	Never
	I have trouble participating in recreational activities with others. . .	**Always**	Usually	Sometimes	Rarely	Never

Appendix 4: Version Changes

Changes in v1.1:

- Changed some language for increased precision (e.g., "pits" to "all-worst health state," "score" to "scoring system").
- Added a motivation for domain selection and a description of item bank construction and scoring.
- Removed speculation about work-in-progress.
- Removed connecting lines in figures with chronic condition impact estimates.

Changes in v1.3:

- We found an error with the method that had created theta estimates for the underlying health states. The scoring function, code, and results were updated to address this error.
- We clarified the domain versions that can be used to generate a PROPr score.
- We were more explicit about how the single attribute functions are combined into a multiattribute function.

Changes in v1.4:

- We found an error with scoring in the Rescaling survey; Figures were updated.